



King's Research Portal

DOI:

[10.1007/s11222-014-9506-9](https://doi.org/10.1007/s11222-014-9506-9)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Nguyen, M. K., Phelps, S., & Ng, W. L. (2015). Simulation based calibration using extended balanced augmented empirical likelihood. *STATISTICS AND COMPUTING*, 25(6), 1093-1112.
<https://doi.org/10.1007/s11222-014-9506-9>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Simulation Based Estimation Using Extended Balanced Augmented Empirical Likelihood

Minh Khoa Ngyuen, Wing Lon Ng and Steve Phelps

Centre for Computational Finance and Economic Agents, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom.

Abstract. This paper introduces an extension of the balanced augmented empirical likelihood method for estimating simulation models. We analyze its performance empirically using Monte-Carlo methods, and demonstrate that our new method increases the flexibility and accuracy of the empirical likelihood approach, while preserving both its limit distribution and its consistency for moment condition models. We illustrate the efficiency of our method in terms of simulation sample size by estimating the parameters of a geometric Brownian motion process.

1. Introduction

[Emerson and Owen \(2009\)](#) recently introduced the balanced augmented empirical likelihood (BAEL), which focuses on the use of the empirical likelihood method for inference about a vector mean. In this paper, we extend the BAEL method and investigate its small sample properties, not only to account for moments of orders higher than one (mean), but also to address more general estimation equations, thus providing more flexibility. Moreover, we propose the extended BAEL (eBAEL) method as a simulation based estimation technique, demonstrating that it provides a consistent estimator. Its efficiency with respect to the simulation sample size is investigated in a benchmark problem.

In many areas of science, models which focus on the properties and behavior of individual components and their interactions — so-called individual-based models, or agent-based models — have become increasingly important ([Bonabeau, 2002](#); [Farmer and Foley, 2009](#); [Shalizi, 2006](#)). These models are often sufficiently complex that deriving closed-form solutions for quantitative aspects of their macroscopic behavior is often impractical if not impossible. Thus these models are often analyzed using Monte-Carlo simulation and empirical methods. However, one criticism of these models is the lack of principled methods for estimating their free parameters against empirically-observed data. To address this problem, this paper introduces an advanced simulation-based estimation procedure based on the empirical likelihood concept.

Former studies in econometrics have proposed several simulation-based estimation techniques, such as simulated maximum likelihood ([Gourieroux and Monfort, 1991](#)), the efficient moment method ([Gallant and Tauchen, 1996](#)), or the indirect inference method ([Gourieroux and Monfort, 1991](#)). However, these estimation procedures face certain difficulties when applied to more complex models. The simulated maximum likelihood method requires that we are able to compute the actual probability density of the likelihood function, and the efficient moment method as well as the indirect inference method suffer the drawback of using an auxiliary model, the latter inducing a source of arbitrariness of capturing the statistical features of the empirical data.

To avoid these shortcomings, we follow a different approach based on the empirical likelihood (EL) framework introduced by [Owen \(1990\)](#). It employs non-parametric likelihood-based tests that can be applied to various functionals of interest such as the mean or the quantiles of a distribution, or regression parameters in multi-sample problems. It is the non-parametric

analogue of the parametric likelihood method and provides efficient estimators and confidence intervals for hypothesis testing. In contrast to the efficient moment method and the indirect inference method, the EL approach does not need an adequate auxiliary model for approximating the likelihood function.

Since the EL method is a non-parametric framework, it has certain advantages over other common simulation based approaches such as the bootstrap. For example, the shape of the EL surface and its confidence region will automatically reflect potential unique features in the observed data giving more emphasis to those parameter values in the model that are best represented by the data (similar to kernel density estimation). Not only are EL methods Barlett correctable (have faster convergence of n^{-2} rather than n^{-1}), but they are also “*range preserving and transformation respecting*” (Hall and La Scara, 1990, p. 110).

Similarly to the common maximum likelihood approach, the idea is to maximize the likelihood of observing particular empirical features as a function of different parameter settings. In fact, the proposed simulation based estimation procedure maximizes the likelihood ratio of the empirical features across simulation outcomes. As the simulation outcomes are determined by the model’s configuration, the proposed simulation based estimator optimizes the empirical likelihood ratio of the empirical moments with respect to the model’s configuration. Because EL ratios can be used for hypothesis tests and confidence regions, the proposed estimation approach can be interpreted as a series of hypothesis tests with a fixed empirically motivated hypothesis and varying simulated data sets, generated from different configurations, similar to a Monte Carlo setting.

The main objective of this paper is to develop an empirical likelihood method for estimating simulation models. Therefore we address the following points. First, in order to provide more flexibility while addressing the convex hull and the under-coverage problem, we consider higher moments and general estimation equations. We demonstrate that our method allows for the construction of confidence intervals, and provides a consistent estimator for moment condition models.

Secondly, in the context of simulation-based estimation, we demonstrate empirically that it converges to the true value and investigate the efficiency of eBAEL in terms of simulation sample size in a simple benchmark problem where we maximize the empirical likelihood of a geometric Brownian motion (GBM) simulator to observe a given empirical moment. Comparing the known optimal analytical results against those obtained by our estimation procedure we show that the procedure yields good small sample estimates. Finally, we also demonstrate that eBAEL provides under certain conditions a consistent simulation based estimator.

This paper provides a range of theoretical and empirical results and will be presented in the following structure. Section 2 introduces the EL approach, in particular the eBAEL for moment condition models and places it in the context of simulation based estimation with the help of a simple example. Section 3 applies the proposed simulation based estimation procedure to a benchmark problem. Section 4 presents all theoretical results. Section 5 concludes. Proofs are provided in the Appendix.

2. Method

This section provides all necessary notation and basic concepts needed for this paper. Section 2.1 introduces the original non-parametric concept of empirical likelihood, while Section 2.2 illustrates hypothesis testing and parameter estimation in the EL framework. Section 2.3 introduces the eBAEL for moment condition models and Section 2.4 puts the latter in the context of simulation based estimation problems.

2.1. Empirical Likelihood

The empirical likelihood function in Owen (1990) is defined as follows.

DEFINITION 1. Given $Y_1, \dots, Y_n \in \mathbb{R}^d$ are i.i.d. with common F . The non-parametric empirical likelihood of any cumulative distribution function F is

$$L(F) = \prod_{i=1}^n [F(y_i) - F(y_i-)], \quad (1)$$

where $F(y-) = P(Y < y)$ and $F(y) = P(Y \leq y)$, thus $P(Y = y) = F(y) - F(y-)$.

REMARK 1. $L(F)$ is the probability of getting exactly the sample values Y_1, \dots, Y_n from F , which reflects exactly the notion of likelihood. To avoid $L(F) = 0$ if F is continuous, a distribution F must place a positive probability on each observed sample moment Y_1, \dots, Y_n in order to have a positive non-parametric likelihood.

The non-parametric likelihood is maximized by the empirical cumulative distribution function (ECDF)

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq y\}}. \quad (2)$$

Furthermore, the empirical likelihood ratio is defined as

$$R(F) = \frac{L(F)}{L(F_n)}. \quad (3)$$

From eq. (3), for a distribution F that places probability w_i on the value Y_i we obtain

$$R(F) = \prod_{i=1}^n \frac{w_i}{\frac{1}{n}}. \quad (4)$$

The numerator is the likelihood of a distribution F with weights on the given observations and the denominator contains the maximum likelihood estimator of all such distributions based on the given observations Y_1, \dots, Y_n .

2.2. Empirical Likelihood: Hypothesis Testing and Parameter Estimation

Suppose we are interested in a parameter $\theta = C(F)$ where $F \in \mathcal{F}$ denotes the set of all distributions that place non-negative weights on the observations $Y_1, \dots, Y_n \in \mathbb{R}^d$. Similarly to eq. (4) we define the profile empirical likelihood ratio function of θ by

$$R(\theta) = \sup \{R(F) \mid C(F) = \theta, F \in \mathcal{F}\}. \quad (5)$$

with

$$\mathcal{F} = \{(w_1, \dots, w_n) \mid w_i \geq 0, \sum_{i=1}^n w_i = 1\}. \quad (6)$$

REMARK 2. Eq. (5) contains $R(F)$. As we can see from eq. (4), this is the ratio between (i) the maximum likelihood estimator of all distributions that place non-negative weights on the given observations (denominator) and (ii) the maximum likelihood estimator of all such distributions, that also satisfy $C(F) = \theta$ (enumerator)

For hypothesis testing we define the profile empirical log likelihood ratio function of θ by

$$W(\theta) = \log R(\theta). \quad (7)$$

It can be used to construct asymptotic confidence intervals (CI) for $\theta_0 = C(F_0)$, the true parameter with respect to the true distribution F_0 of the observations Y_1, \dots, Y_n . For certain parameters and under some regularity conditions (Owen, 1990; Qin and Lawless, 1994), it has been demonstrated that

$$-2W(\theta_0) \rightarrow \chi^2 \quad (8)$$

as $n \rightarrow \infty$, and the $100(1 - \alpha)\%$ CI is

$$\{\theta : -2W(\theta) \leq \chi^2(1 - \alpha)\}, \quad (9)$$

where $\chi^2(1 - \alpha)$ is the $(1 - \alpha)^{th}$ quantile of the χ^2 -distribution. On the other hand, Qin and Lawless (1994) also demonstrated that for given observations Y_1, \dots, Y_n the empirical maximum likelihood of estimate of θ_0 is given by

$$\hat{\theta}_{EL} = \underset{\theta}{argmax} [W(\theta)]. \quad (10)$$

In the following, we contribute to the literature by further extending the Balanced Augmented EL (BAEL) approach in Emerson and Owen (2009). We address two particular problems: first the convex hull problem (i.e. the EL function is not always defined), second the under-coverage problem.

The under-coverage problem refers to the observation of DiCiccio et al. (1991) that for small samples the actual coverage probability of EL confidence intervals (CI) are smaller than the nominal ones. The coverage error is a direct consequence of the EL ratio statistic which is used to construct CIs, and converges only asymptotically. These issues are addressed by ? and Emerson and Owen (2009) who suggest augmenting the data set with artificial data points and balancing possible under-coverage in the original data set giving rise to *Balanced Augmented Empirical Likelihood* (BAEL). For moments of order one, i.e. the mean, BAEL yields superior small sample properties.

2.3. Extended Balanced Augmented Empirical Likelihood for Estimation Equations

In this section we describe our new methodology, extended Balanced Augmented Empirical Likelihood which we abbreviate eBAEL. eBAEL is an approach for hypothesis testing and estimation of moment condition models.

Suppose we are interested in a q -dimensional parameter $\theta = C(F)$, for some function C of some distribution F . Moreover, additional information on θ and F is available in the form of $l \geq q$ functionally independent unbiased estimation equations $g_j(Y, \theta)$, $j = 1, \dots, l$ or in vector form:

$$g(Y, \theta) = (g_1(Y, \theta), \dots, g_l(Y, \theta))', \quad (11)$$

where Y has distribution F . Now suppose some sample Y_1, \dots, Y_n is i.i.d. with unknown distribution F_0 , then θ_0 is the true solution of the moment condition model such that

$$E_{F_0}[g(Y, \theta_0)] = 0. \quad (12)$$

Let g_i be the short version of $g(Y_i, \theta)$ for Y_i , $i = 1, \dots, n$. Then the eBAEL profile empirical likelihood function of θ and sample Y_1, \dots, Y_n is defined as

$$\tilde{R}(\theta) = \sup \left\{ \prod_{i=1}^{n+2} \frac{w_i}{\frac{1}{n+2}} \mid \sum_{i=1}^{n+2} w_i g_i = 0, \sum_{i=1}^{n+2} w_i = 1, w_i \geq 0 \right\}. \quad (13)$$

The profile empirical log-likelihood function is given by

$$\tilde{W}(\theta) = \log \tilde{R}(\theta). \quad (14)$$

The points g_{n+1} and g_{n+2} are two new sample points around the mean

$$\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g_i \quad (15)$$

in direction u

$$u = \frac{\bar{g}_n - 0}{\|\bar{g}_n - 0\|}, \quad (16)$$

where $\|\cdot\|$ is a vector norm. eBAEL adds a point nearer to the zero vector when the uncertainty is smaller in that direction u , and further away when the uncertainty is larger in that direction. Let \hat{S} denote the sample covariance matrix

$$\hat{S} = \frac{1}{n-1} \sum_{i=1}^n (g_i - \bar{g}_n)(g_i - \bar{g}_n)'. \quad (17)$$

Then

$$c_u = \left(u' \hat{S}^{-1} u \right)^{-\frac{1}{2}} \quad (18)$$

is the inverse Mahalanobis distance of a unit vector from \bar{g}_n in the direction of u . For a fixed $s \in \mathbb{R}$, g_{n+1} and g_{n+2} are defined by

$$g_{n+1} = -s c_u u \quad (19)$$

and

$$g_{n+2} = 2\bar{g}_n + s c_u u. \quad (20)$$

That is, we place a new point near the zero vector when the covariance \tilde{S} in direction u is smaller and further away when the covariance in that direction is larger, thereby insuring that the zero vector is included in the convex hull of $\{g(Y_i, \theta), i = 1, \dots, n\}$.

REMARK 3. Eq. (13) contains the ratio between (i) the maximum likelihood estimator of all distributions that place non-negative weights on the given observations (denominator) and (ii) the maximum likelihood estimator of all such distributions, that also satisfy the moment condition model with parameter θ (enumerator).

REMARK 4. Without augmenting the data set with these two new points, 0 might not be in the convex hull of $\{g(Y_i, \theta), i = 1, \dots, n\}$ and as a consequence eq. (13) would have no solution and be undefined. Moreover, by including the second point g_{n+2} the original sample mean is maintained since $\frac{1}{n+2} \sum_{i=1}^{n+2} g_i = \bar{g}_n$.

The explicit expressions for $\tilde{R}(\theta)$ and $\tilde{W}(\theta)$ are derived by using Lagrange multipliers:

$$\mathcal{L} = \sum_{i=1}^{n+2} \log(w_i) + \tau \left(1 - \sum_{i=1}^{n+2} w_i \right) - (n+2) \lambda' \left(\sum_{i=1}^{n+2} w_i g_i(\theta) \right) + (n+2) \log(n+2).$$

With the first order condition for \mathcal{L}

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{1}{w_i} - \tau - (n+2) \lambda' g_i(\theta) = 0 \quad (21)$$

we get

$$\sum_{i=1}^{n+2} w_i \frac{\partial \mathcal{L}}{\partial w_i} = (n+2) - \tau = 0$$

which gives us $\tau = n+2$. Substituting this in eq. (21), the optimal weights w_i^* , $i = 1, \dots, n+2$ are given by

$$w_i^* = \frac{1}{(n+2)(1 + \lambda' g_i(\theta))}, \quad (22)$$

where with the condition $\sum_{i=1}^{n+2} w_i g_i(\theta) = 0$, λ must satisfy

$$\frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i(\theta)}{1 + \lambda' g_i(\theta)} = 0. \quad (23)$$

Therefore $\tilde{W}(\theta)$ can be written as

$$\tilde{W}(\theta) = \sum_{i=1}^{n+2} \log((n+2) w_i^*) = - \sum_{i=1}^{n+2} \log(1 + \lambda' g_i(\theta)). \quad (24)$$

Similarly to eq. 10 in Section 2.2, the eBAEL estimator of θ_0 is

$$\hat{\theta}_{eBAEL} = \underset{\theta}{argmax} [\tilde{W}(\theta)] \quad (25)$$

and $\tilde{W}(\theta)$ can be used to construct asymptotic confidence intervals as

$$-2\tilde{W}(\theta_0) \rightarrow \chi^2. \quad (26)$$

The latter is shown in Section 4. Moreover, we show that eBAEL also provides a consistent estimator of θ_0 . In the following Lemma we provide with same arguments as in [Qin and Lawless \(1994\)](#) (see p. 304 and p. 317) that $\tilde{W}(\theta)$ is continuous in θ_0 .

LEMMA 1. *If $g(Y, \theta)$ is continuous in the neighborhood of θ_0 and $\Sigma_g(\theta_0) < \infty$, then $\tilde{W}(\theta)$ is continuous in θ_0 .*

PROOF. See Section A.1

Algorithm 1: Compute $\tilde{W}_\gamma(\hat{\mu}^e)$

Data: $\hat{\mu}^e, \gamma$
Result: $\tilde{W}_\gamma(\hat{\mu}^e)$
begin
 1. Generate augmented simulation sample:
 for $j \leftarrow 1$ **to** $n + 2$ **do**
 if $j \leq n$ **then**
 | generate simulation data $y_i^s(\gamma)$
 else
 | generate additional sample points:
 $y_{n+1}^s(\gamma) = \hat{\mu}^e - sc_u u;$ (see eq. (19))
 $y_{n+2}^s(\gamma) = 2\bar{y}_n(\gamma) - \hat{\mu}^e + sc_u u;$ (see eq. (20))
 end
end
 2. Compute the root of $f(\lambda) := \sum_{i=1}^{n+2} \frac{y_i^s(\gamma) - \hat{\mu}^e}{1 + \lambda'(y_i^s(\gamma) - \hat{\mu}^e)}$: (see eq. (23))
 $\lambda_0 = 0; \epsilon = 10^{-6}$
 while $f(\lambda_k) > \epsilon$ **do**
 | $\lambda_{k+1} = \lambda_k - \frac{f(\lambda_k)}{f'(\lambda_k)}$
end
 $\lambda = \lambda_k$.
 3. Compute the root of value of $\tilde{W}_\gamma(\hat{\mu}^e)$:
 $\tilde{W}_\gamma(\hat{\mu}^e) = -\sum_{i=1}^{n+2} \log\{1 + \lambda'(y_i^s(\gamma) - \hat{\mu}^e)\}$ (see eq. (24))
end

2.4. Simulation Based Estimation and eBAEL: An Example

In this section we demonstrate the problem of simulation based estimation with an example and discuss the eBAEL approach and contrast it with the simulated moment method (SMM, see [McFadden \(1989\)](#)). Suppose the following simple structural model

$$E[f(Y, \gamma)] = \mu(\gamma), \quad (27)$$

with structural parameter $\gamma \in \Gamma$, f a moment function and m empirical observations $\mathbf{y}^e = (y_1^e, \dots, y_m^e)$. Moreover suppose there exists γ_0 , such that γ_0 is the true parameter of eq. (27) given the observations \mathbf{y}^e . In case $\mu(\gamma)$ is not analytically tractable it can be replaced by its simulated counterpart

$$\tilde{\mu}(\gamma) = \frac{1}{n} \sum_{i=1}^n f(y_i^s(\gamma)), \quad (28)$$

where $\tilde{\mu}(\gamma)$ is computed from n simulation samples $\mathbf{y}^s = (y_1^s(\gamma), \dots, y_n^s(\gamma))$.

For the case where $\mu(\gamma)$ is analytically tractable, the generalised method of moments (GMM, see [Hansen and Singleton \(1982\)](#)) estimate of the parameter γ_0 is given by

$$\begin{aligned} \hat{\gamma}_m^{GMM} &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} [\hat{\mu}^e - \mu(\gamma)]' D [\hat{\mu}^e - \mu(\gamma)] \\ &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} \left[\frac{1}{m} \sum_{j=1}^m f(y_j^e) - \mu(\gamma) \right]' D \left[\frac{1}{m} \sum_{j=1}^m f(y_j^e) - \mu(\gamma) \right], \end{aligned} \quad (29)$$

where D is a weight matrix (see also Hansen and Singleton (1982)). For the case when $\boldsymbol{\mu}(\gamma)$ is not analytical tractable, the SMM methodology replaces the latter with its simulated counterpart $\tilde{\boldsymbol{\mu}}(\gamma)$ (see eq. (28)) and the SMM estimate of γ_0 is given by

$$\hat{\gamma}_{m,n}^{SMM} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} [\hat{\boldsymbol{\mu}}^e - \tilde{\boldsymbol{\mu}}(\gamma)]' D [\hat{\boldsymbol{\mu}}^e - \tilde{\boldsymbol{\mu}}(\gamma)] \quad (30)$$

with

$$\hat{\boldsymbol{\mu}}^e - \tilde{\boldsymbol{\mu}}(\gamma) = \frac{1}{m} \sum_{j=1}^m f(y_j^e) - \frac{1}{n} \sum_{i=1}^n f(y_i^s(\gamma)). \quad (31)$$

Here both the GMM as the SMM estimates are given by minimizing the weighted squared error between the empirical and the model's unconditional moment. As an alternative this paper proposes the EL framework for simulation based estimation problems as in eq. (27). The idea is to evaluate the fitness of the model not in terms of error but in terms of likelihood. Given a parameter γ and the corresponding simulation sample $\mathbf{y}_\gamma^s = (y_1^s(\gamma), \dots, y_n^s(\gamma))$ and the structural model in eq. (27), the EL approach determines the fitness by calculating the likelihood. In particular using eBAEL in Section 2.3 the proposed estimate of γ_0 is given by

$$\hat{\gamma}_{m,n}^{eBAEL} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} [\tilde{W}_\gamma(\hat{\boldsymbol{\mu}}^e)], \quad (32)$$

where $\tilde{W}_\gamma(\hat{\boldsymbol{\mu}}^e) = \log(\tilde{R}_\gamma(\hat{\boldsymbol{\mu}}^e))$ with

$$\tilde{R}_\gamma(\hat{\boldsymbol{\mu}}^e) = \sup \left\{ \prod_{i=1}^{n+2} (n+2) w_i \mid \sum_{i=1}^{n+2} w_i f(y_i^s(\gamma)) = \hat{\boldsymbol{\mu}}^e, \sum_{i=1}^{n+2} w_i = 1, w_i \geq 0 \right\}. \quad (33)$$

This estimate maximizes the empirical likelihood ratio for $\hat{\boldsymbol{\mu}}^e$ and simulation samples \mathbf{y}_γ^s with $\gamma \in \Gamma$. Algorithm 1 presents the pseudo code for computing $\tilde{W}_\gamma(\hat{\boldsymbol{\mu}}^e)$ in eq. (32) using Remark 2.6 (see also eq. (48) in Section 3).

In order to address a more general setting than the mean $\hat{\boldsymbol{\mu}}^e$, we define the following notation. Let γ be the parameter of any structural model with a domain Λ and $\mathbf{y}_\gamma^s = (y_1^s(\gamma), \dots, y_n^s(\gamma))$ denotes the simulation sample i.i.d. with some unknown distributions $F(\gamma)$. Furthermore let θ be any quantity that we like to maximize the likelihood given the simulation outcomes $\{\mathbf{y}_\gamma^s\}_{\gamma \in \Lambda}$ such that θ is a function of the unknown distributions F with some unbiased estimation equation g , i.e.

$$E[g(Y, \theta)] = 0.$$

Then the estimator in eq. (32) becomes

$$\hat{\gamma}_{m,n}^{eBAEL} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} [\tilde{W}_\gamma(\theta)].$$

As this estimation procedure estimates against some empirical quantity, the precise notation is

$$\hat{\gamma}_{m,n}^{eBAEL} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} [\tilde{W}_\gamma(\hat{\theta}_m)], \quad (34)$$

where $\tilde{W}_\gamma(\hat{\theta}_m) = \log(\tilde{R}_\gamma(\hat{\theta}_m))$ with

$$\tilde{R}_\gamma(\hat{\theta}_m) = \sup \left\{ \prod_{i=1}^{n+2} (n+2) w_i \mid \sum_{i=1}^{n+2} w_i g_i = 0, \sum_{i=1}^{n+2} w_i = 1, w_i \geq 0 \right\} \quad (35)$$

and $g_i = g_i(y_i^s(\gamma), \hat{\theta}_m)$.

The subscripts of $\hat{\gamma}_{m,n}^{eBAEL}$ emphasize the number of empirical observations and the number of simulated samples used for calculating the empirical likelihood underlying the estimate $\hat{\gamma}_{m,n}^{eBAEL}$.

REMARK 5. Setting $\hat{\theta}_m = \hat{\mu}^e$ and $g_i(y_i^s(\gamma), \hat{\mu}^e) = f(y_i^s(\gamma)) - \hat{\mu}^e$ in eq. (35) we get the same result as in eq. (33).

REMARK 6. In the case of simulation based estimation the parameter of interest is the structural parameter γ , which should not be confused with θ the parameter for moment condition models in Section 2.3. In this context θ is fixed and given by its empirical counterpart $\hat{\theta}_m$.

However, for a true simulation based estimation methodology, the estimator in eq. (34) needs to be consistent, similar to the SMM methodology (Pakes and Pollard, 1989). That is, despite the noise from empirical and simulated data, the estimator must converge to the true parameter of the estimated model with respect to the empirical data.

Suppose the empirical observations $\mathbf{y}^e = (y_1^e, \dots, y_m^e)$ underlying $\hat{\theta}_m$ are i.i.d. with common F_0 , then there exists θ_0 such that

$$E_{F_0}[g(Y, \theta_0)] = 0. \quad (36)$$

Similarly, suppose that, for all $\gamma \in \Gamma$, (a) the simulation samples $\mathbf{y}^s(\gamma) = (y_1^s(\gamma), \dots, y_n^s(\gamma))$ are i.i.d. with $F(\gamma)$, and (b) $\theta(\gamma)$ is the true value of

$$E_{F(\gamma)}[g(Y, \theta(\gamma))] = 0. \quad (37)$$

Then true parameter γ_0 is given by

$$\theta(\gamma_0) = \theta_0. \quad (38)$$

By definition the estimate in eq. (34) is a consistent estimator of γ_0 if

$$\hat{\gamma}_{m,n}^{eBAEL} \xrightarrow{P} \gamma_0 \quad (39)$$

as $m, n \rightarrow \infty$. In the next section we empirically demonstrate its ability to converge to the true parameter γ_0 in a simulation study. The consistency of eBAEL as simulation based estimator is provided in Section 4.2.

3. Application to Geometric Brownian Motion (GBM)

In this section we illustrate the application of our simulation based estimation approach in Section 2.4 to a simple benchmark problem, viz. estimating the parameters of a geometric Brownian motion process, and we analyze its performance in comparison to AEL (?) and SMM (McFadden, 1989; ?). We demonstrate (in Section 3.1) that our method is able to estimate to the true underlying parameters of the GBM process, and (in Section 3.2) whether it is able to do so with a small number of samples (efficiency). This is particularly important for estimation of complex models (e.g., agent-based models) in which each sample corresponds to the execution of a simulation which may be very costly in terms of CPU time.

The GBM is given by

$$S_t = S_0 e^{X_t}, \quad (40)$$

where $X_t = (\alpha - \frac{\delta^2}{2})t + \delta B_t$, and B_t is a standard Brownian Motion, and the (log-) increments

$$r_t = \log(S_t) - \log(S_{t-\Delta t}) \quad (41)$$

over an interval Δt are normally distributed with

$$N(\mu, \sigma^2) = N\left(\left(\alpha - \frac{\delta^2}{2}\right) \Delta t, \delta^2 \Delta t\right). \quad (42)$$

As the GBM is entirely governed by α and δ , the (structural) parameter of the GBM is given by

$$\gamma = (\alpha, \delta), \quad (43)$$

which is our parameter of interest and to be estimated.

For a given empirical time series with $p = T/\Delta t$ increments $\{\hat{r}_t\}_{t=1,\dots,p}$ the sample mean and variance is given by

$$\bar{r}^e = \frac{1}{p} \sum_{t=1}^p \hat{r}_t, \quad (44)$$

$$\hat{s}^e = \frac{1}{p-1} \sum_{t=1}^p (\hat{r}_t - \bar{r}^e)^2. \quad (45)$$

Now suppose there are m empirical time series, giving the following empirical observations $\{y_j^e = (\bar{r}_j^e, s_j^e), j = 1, \dots, m\}$. Furthermore, the simulated GBM counterparts are given by $y_i^s(\gamma) = (\bar{r}_i(\gamma), s_i^2(\gamma)), i = 1, \dots, n$, where $y_i^s(\gamma)$ are computed from the simulated GBM increments $\{r_i^s(\gamma)\}_{t=1,\dots,p}$. In order to emphasize the empirical or simulated nature of the variables m and n , we will use in this section m_{emp} and n_{sim} .

Similarly to eq. (30), with f the identity mapping, the SMM estimate of the GBM is for example given by minimizing the weighted squared error between the average empirical and the GBM observations

$$\hat{\gamma}_{m_{emp}, n_{sim}}^{SMM} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} [\hat{\mu}^e - \tilde{\mu}(\gamma)]' D [\hat{\mu}^e - \tilde{\mu}(\gamma)]$$

with

$$\hat{\mu}^e - \tilde{\mu}(\gamma) = \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} y_j^e - \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} y_i^s(\gamma) \quad (46)$$

$$= \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} \begin{pmatrix} \bar{r}_j^e \\ \hat{s}_j^e \end{pmatrix} - \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \begin{pmatrix} \bar{r}_i(\gamma) \\ s_i^2(\gamma) \end{pmatrix} \quad (47)$$

Similarly to eqs. (32) and (33) the corresponding eBAEL estimator is given by

$$\hat{\gamma}_{m_{emp}, n_{sim}}^{eBAEL} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} [\tilde{W}_\gamma(\hat{\mu}^e)], \quad (48)$$

with

$$\tilde{R}_\gamma(\hat{\mu}^e) = \sup \left\{ \prod_{i=1}^{n_{sim}+2} (n_{sim} + 2) w_i \mid \sum_{i=1}^{n_{sim}+2} w_i y_i^s(\gamma) = \hat{\mu}^e, \sum_{i=1}^{n_{sim}+2} w_i = 1, w_i \geq 0 \right\}. \quad (49)$$

Table 1. Pseudo Empirical Observations.

Sample size m_{emp}	$\hat{\mu}_{m_{emp}}^e(1)$	$\hat{\mu}_{m_{emp}}^e(2)$
250	0.015650217	0.010121132
300	0.015280075	0.010106784
350	0.015253495	0.010077696
400	0.015155546	0.010053884
500	0.014950879	0.010023655

3.1. Convergence

In order to demonstrate the convergence of our approach we compute the EL estimate as in eq. (48) when estimating against some pseudo empirical moments generated from a GBM with a fixed configuration of the “true” parameter $\gamma_0 = (\alpha_0, \delta_0) = (2, 1)$.

The pseudo-empirical moments

$$\hat{\mu}_{m_{emp}}^e = \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} y_j^e(\gamma_0) = \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} \begin{pmatrix} \bar{r}_j^e(\gamma_0) \\ \hat{s}_j^e(\gamma_0) \end{pmatrix} \quad (50)$$

are given in Table 1. These are simply the average of sample means and variances of the increments; that is, the average of m_{emp} sample means and variances from m_{emp} GBM time series with $\Delta t = 1/100$, $T = 1$, where $\hat{\mu}_{m_{emp}}^e(1) = \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} \bar{r}_j^e$ and $\hat{\mu}_{m_{emp}}^e(2) = \frac{1}{m_{emp}} \sum_{j=1}^{m_{emp}} \hat{s}_j^e$.

As we increase the (pseudo) empirical sample size m_{emp} and the simulation sample size n_{sim} , our estimate $\hat{\gamma}_{m_{emp}, n_{sim}}^{eBAEL}$ in eq. (48) must converge to $\gamma_0 = (2, 1)$. Indeed, this is illustrated in Figures 1 and 2, displaying the mean square errors (MSE) of 500 estimates $\hat{\gamma}_{m_{emp}, n_{sim}}^{eBAEL}$, whereas Figures 3 to 6 display the corresponding standard errors of the estimates, while using either $n_{sim} = 100$ or $n_{sim} = 500$ GBM simulations. These graphs show that as we increase the empirical sample size m_{emp} and the simulation sample size n_{sim} we obtain more accurate (in terms of estimation variability) estimates as $\hat{\gamma}_{m_{emp}, n_{sim}}^{eBAEL}$ tending towards the true configuration $\gamma_0 = (2, 1)$ by decreasing mean square estimation errors as well as standard errors.

3.2. Efficiency

In this efficiency experiment we consider the small sample performance (with respect to the simulation sample size n) of our estimate in comparison to AEL and SMM when estimating the GBM against $\hat{\mu}^e = (\bar{r}^e, \hat{s}^e) = (0.015, 0.01)$. Furthermore for this experiment we set $\Delta t = 1/100$ and $T = 1$ when simulating the GBM process. Assuming there is no noise from empirical data, that is $\hat{\mu}^e = \mu_0 = (0.015, 0.01)$ with eq. (42) it follows that the true GBM parameter setting for this experiment is also given by $\gamma_0 = (\alpha_0, \delta_0) = (2, 1)$. A sample surface of $\tilde{W}(\hat{\mu}^e)$ for the given experiment over a mesh of (α, δ) is depicted in Figure 7.

For our small sample experiment we have executed the described GBM estimation a 1000 times, using the SMM estimate in eq. (46) and our EL estimate in eq. (48) with $\hat{\mu}^e = (0.015, 0.01)$. The AEL type estimate is similar to the our estimate and is simply given by

$$\hat{\gamma}^{AEL} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} [W_{\gamma}^*(\hat{\mu}^e)], \quad (51)$$

where

$$R_{\gamma}^*(\hat{\mu}^e) = \sup \left\{ \prod_{i=1}^{n_{sim}+1} (n_{sim} + 1) w_i \mid \sum_{i=1}^{n_{sim}+1} w_i y_i^s(\gamma) = \hat{\mu}^e, \sum_{i=1}^{n_{sim}+1} w_i = 1, w_i \geq 0 \right\}. \quad (52)$$

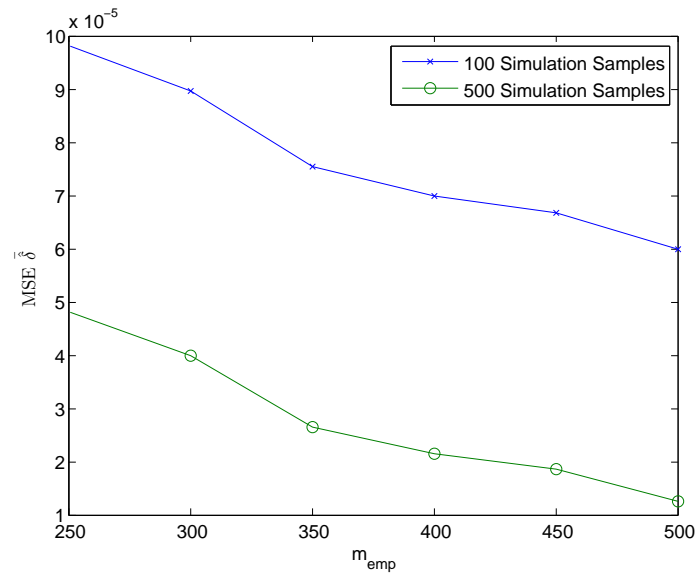


Figure 1. Mean Squared Error (MSE) of $\hat{\delta}$ against m_{emp} the number of empirical time series.

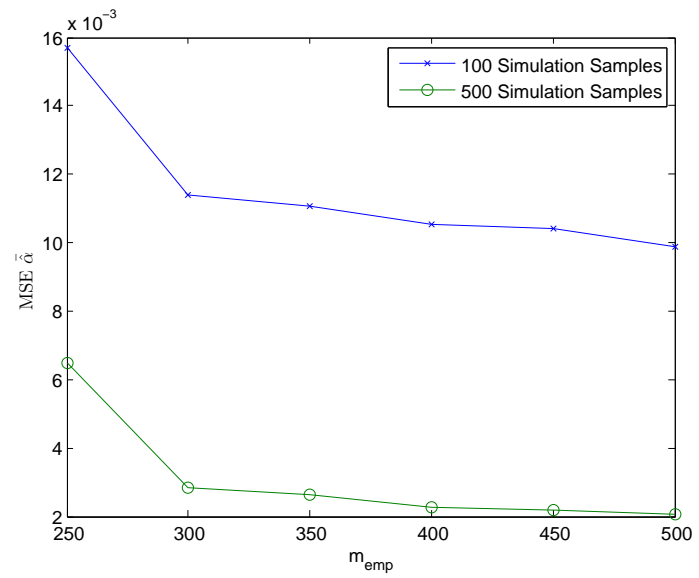


Figure 2. Mean Squared Error (MSE) of $\hat{\alpha}$ against m_{emp} the number of empirical time series.

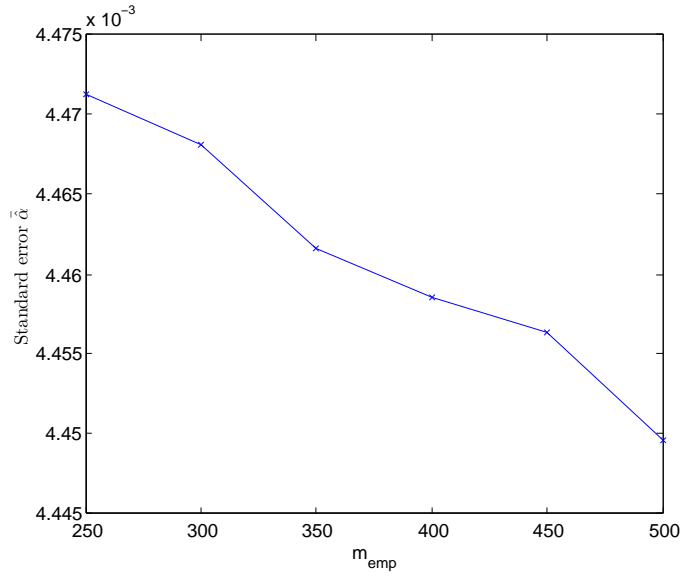


Figure 3. Standard error of $500 \hat{\alpha}$ against m_{emp} the number of empirical time series, while using $n_{sim} = 100$ GBM simulation samples

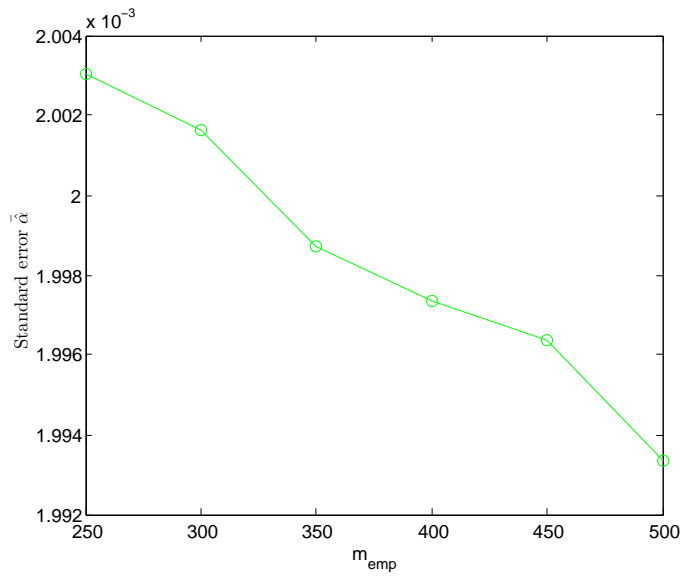


Figure 4. Standard error of $500 \hat{\alpha}$ against m_{emp} the number of empirical time series, while using $n_{sim} = 500$ GBM simulation samples

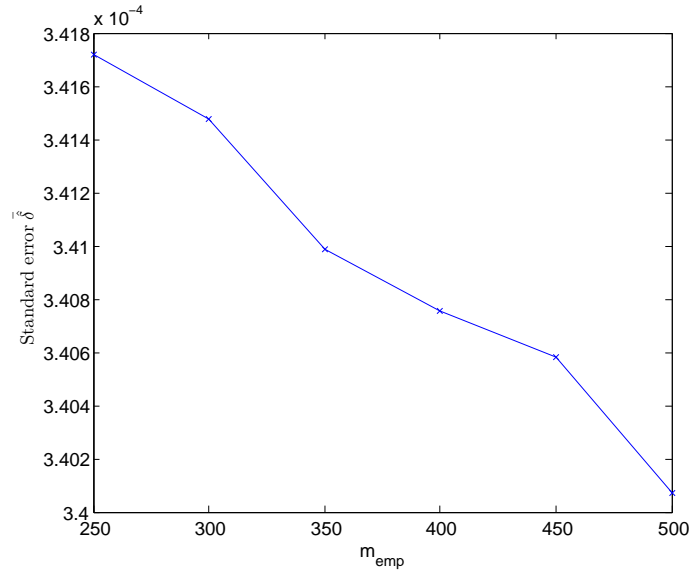


Figure 5. Standard error of $500 \hat{\delta}$ against m_{emp} the number of empirical time series, while using $n_{sim} = 100$ GBM simulation samples

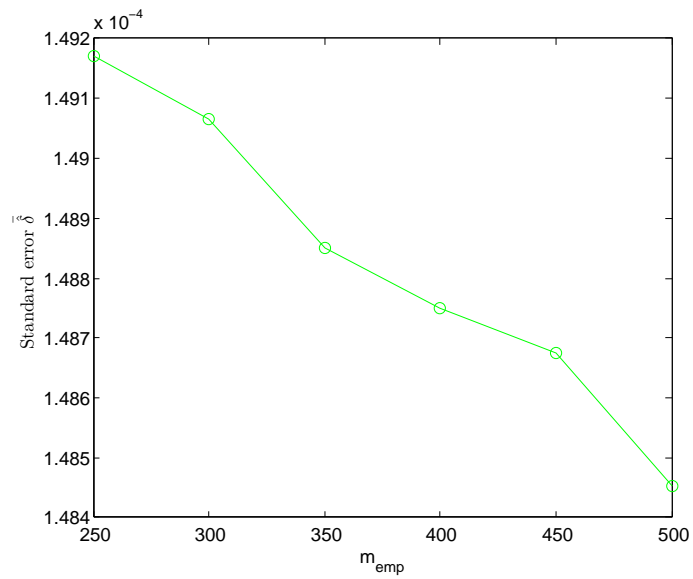


Figure 6. Standard error of $500 \hat{\delta}$ against m_{emp} the number of empirical time series, while using $n_{sim} = 500$ GBM simulation samples

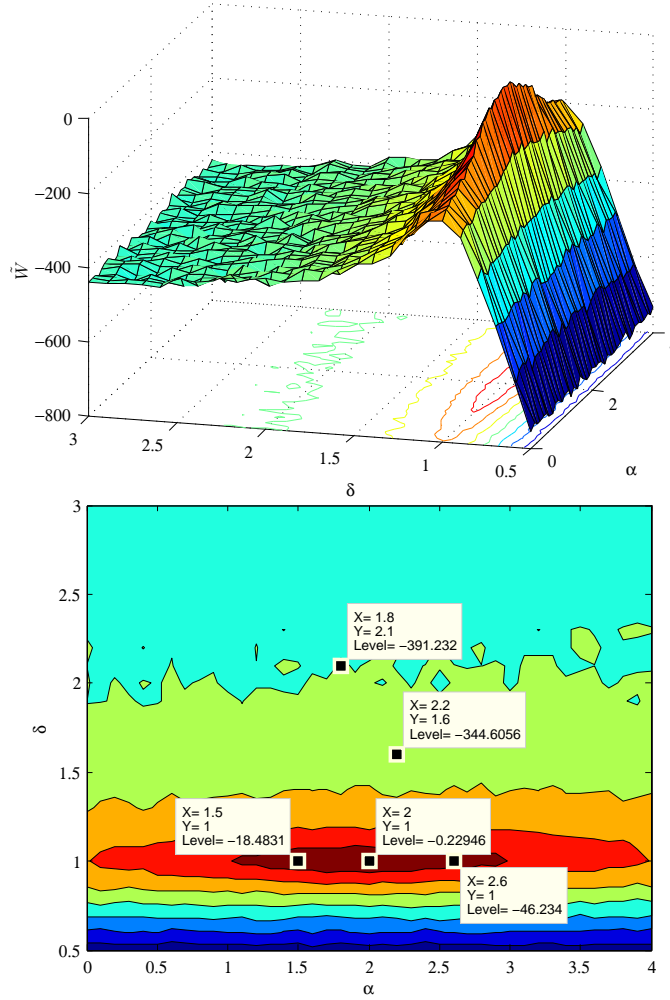


Figure 7. Log Empirical Likelihood Ratio. The top panel displays the log empirical likelihood ratio surface over a grid of the GBM parameters α and δ . For each configuration 300 sample paths are simulated from which 300 means and variances are sampled. For the given moment sample we compute the log empirical likelihood ratio for $\hat{\mu}_e = (0.015, 0.01)$. The bottom panel shows the corresponding contour plot of the log empirical likelihood ratio surface.

Table 2. Comparing eBAEL against AEL and SMM. This table reports the MSE as well as the variance of the estimates from 1000 executions of estimating the GBM against the moment hypothesis $y_e = (0.015, 0.01) = (\mu_e, \sigma_e^2)$ using eBAEL, AEL and the SMM method.

number of samples (n_{sim})	eBAEL		AEL		SMM	
	MSE($\hat{\alpha}$)	MSE($\hat{\delta}$)	MSE($\hat{\alpha}$)	MSE($\hat{\delta}$)	MSE($\hat{\alpha}$)	MSE($\hat{\delta}$)
15	0.0645	0.0003	0.0652	0.0004	0.0677	0.0125
20	0.0477	0.0003	0.0498	0.0003	0.0522	0.0084
30	0.0321	0.0002	0.0319	0.0002	0.0334	0.0002
50	0.0187	0.0001	0.0188	0.0001	0.0196	0.0001

number of samples	eBAEL		AEL		SMM	
	Var($\hat{\alpha}$)	Var($\hat{\delta}$)	Var($\hat{\alpha}$)	Var($\hat{\delta}$)	Var($\hat{\alpha}$)	Var($\hat{\delta}$)
15	0.0645	0.0003	0.0652	0.0004	0.0678	0.0124
20	0.0477	0.0003	0.0498	0.0003	0.0522	0.0084
30	0.0322	0.0002	0.0319	0.0002	0.0334	0.0002
50	0.0187	0.0001	0.0188	0.0001	0.0196	0.0001

with an artificial extra point

$$y_{n_{sim}+1}^s(\gamma) = \hat{\mu}^e - \frac{1}{2} \log(n_{sim}) [\bar{y}_{n_{sim}}(\gamma) - \hat{\mu}^e]. \quad (53)$$

Table 2 reports the the mean squared error (MSE) as well as the variance of the estimates from those 1000 executions. For each individual estimation we employ the common random numbers technique: that is, within an estimation all n_{sim} sample paths of the stochastic process for any arbitrary parameter setting γ will be generated from the same stream of random numbers. However, for different estimations, the fixed stream of random numbers will differ. The results, presented in Table 2, show that all estimates are getting naturally more accurate with increasing sample numbers. The results also demonstrate that the MSE and the variance of the EL based estimates are better than the corresponding SMM figures in particular for small samples $n_{sim} = 15, 20, 30$. Moreover, within the EL based methods, the eBAEL outperforms the AEL approach in particular for very small samples ($n_{sim} = 15, 20$).

4. Theoretical Results

This Section presents the theoretical results of this paper. In Section 4.1 we demonstrate that for moment condition models the eBAEL ratio has for θ_0 a χ^2 limit distribution, allowing us to construct asymptotic CIs, and otherwise it diverges. Moreover, it provides a consistent estimator of θ_0 . Section 4.2 addresses the consistency of eBAEL as a simulation based estimator.

4.1. eBAEL and Moment Condition Models

The first two Theorems provide the limiting behavior of \tilde{W} as $n \rightarrow \infty$. The first one addresses the case when $\theta = \theta_0$ and the second one $\theta \neq \theta_0$.

THEOREM 1. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ iid with some unknown distribution F_0 . For $\theta \in \Theta \subseteq \mathbb{R}^q$ and $Y \in \mathbb{R}^d$, let $g(Y, \theta) \subseteq \mathbb{R}^l$ with $l \geq q$. Moreover let $\theta_0 \in \Theta$ such that $E_{F_0}[g(Y, \theta_0)] = 0$ and $\sum_g = \text{Var}_{F_0}[g(Y, \theta_0)] < \infty$ with rank q . Then as $n \rightarrow \infty$ and for a fixed value $s \in \mathbb{R}$ we have*

$$-2\tilde{W}(\theta_0) \xrightarrow{d} \chi_q^2$$

as $n \rightarrow \infty$.

PROOF. See A.2.

The following Theorem establishes the asymptotic behavior of \tilde{W} when $\theta \neq \theta_0$.

THEOREM 2. *Let $Y_1, \dots, Y_n \in \mathbb{R}^d$ iid with some unknown distribution F_0 . For $\theta \in \Theta \subseteq \mathbb{R}^q$ and $Y \in \mathbb{R}^d$, let $g(Y, \theta) \in \mathbb{R}^l$ with $l \geq q$. Let $\theta_0 \in \Theta$ such that $E_{F_0}[g(Y, \theta_0)] = 0$ then for $\theta \neq \theta_0$ where*

$$\|E_{F_0}[g(Y, \theta)]\| > 0$$

and suppose

$$\Sigma_g(\theta) = E[(g(Y, \theta) - E[g(Y, \theta)])(g(Y, \theta) - E[g(Y, \theta)])'] < \infty.$$

Then for a fixed value $s \in \mathbb{R}$ we have $-n^{-1/3}\tilde{W}(\theta) \rightarrow \infty$ in probability as $n \rightarrow \infty$.

PROOF. See A.3

REMARK 1. *Suppose $\theta_0 \in \Theta \subseteq \mathbb{R}^q$ and $\Sigma_g(\theta) < \infty$ for all $\theta \in \Theta$, then with Theorem 2 it follows $-n^{-1/3}\tilde{W}(\theta) \rightarrow \infty$ in probability for all $\theta \in \Theta$ with $\theta \neq \theta_0$. Now suppose as before $\Sigma_g(\theta) < \infty$ for all $\theta \in \Theta$, but either $\theta_0 \notin \Theta$ or it doesn't exist then by definition $\|E_{F_0}[g(Y, \theta)]\| > 0$ for all $\theta \in \Theta$, hence with Theorem 2 it follows $-n^{-1/3}\tilde{W}(\theta) \rightarrow \infty$ in probability for all $\theta \in \Theta$.*

The next theorem shows that the eBAEL approach provides us with

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \tilde{W}(\theta)$$

which is a consistent estimator of θ_0 given the following assumptions.

ASSUMPTION 1.

- (a) $\theta_0 \in \Theta$ is the unique solution to $E[g(y, \theta)] = 0$.
- (b) Θ is compact.
- (c) $g(y, \theta)$ is continuous at each $\theta \in \Theta$ with probability 1.
- (d) $E\left[\sup_{\theta \in \Theta} \|g(y, \theta)\|^\alpha\right] < \infty$ for some $\alpha > 2$.

Following Newey and Smith (2004), the EL estimator is the solution to the saddle point problem and can be written as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \max_{\lambda \in \hat{\Lambda}_n(\theta)} \sum_{i=1}^{n+2} \rho(\lambda' g_i(\theta)), \quad (54)$$

where $\hat{\Lambda}_n(\theta) = \{\lambda : \lambda' g_i(\theta) \in \mathfrak{S}, i = 1, \dots, n+2\}$ with $\mathfrak{S} = (-\infty, 1)$ and $\rho(v) = \log(1-v)$. Let $\rho_j(v) = \partial^j \rho(v) / \partial v^j$ and $\rho_j = \rho_j(0)$, then we have $\rho_2(v) < 0$ for all $v \in \mathfrak{S}$ and in particular $\rho_1 = \rho_2 = -1$. Furthermore we use the following notation $\hat{g}(\hat{\theta}) = \frac{1}{n+2} \sum_{i=1}^{n+2} g_i(\hat{\theta})$.

THEOREM 3. *If Assumption 1 is satisfied, then $\hat{\theta} \rightarrow \theta_0$ in probability as $n \rightarrow \infty$, $\hat{g}(\hat{\theta}) = O_p(n^{-1/2})$, $\hat{\lambda} = \underset{\lambda \in \hat{\Lambda}_n(\hat{\theta})}{\operatorname{argmax}} \sum_{i=1}^{n+1} \rho(\lambda' g_i(\hat{\theta})) / (n+2)$ exists with probability approaching 1 and $\hat{\lambda} = O_p(n^{-1/2})$.*

PROOF. See A.4.

4.2. eBAEL a Consistent Simulation Based Estimator

In this section we demonstrate that eBAEL provides a consistent simulation based estimator. As in Section 2.4 the simulation based estimator is given by

$$\hat{\gamma}_{m,n} = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \left[\tilde{W}_{\gamma} \left(\hat{\theta}_m \right) \right] \quad (55)$$

and as before θ_0 is the true value of $\hat{\theta}_m$ given by empirical data $\mathbf{y}^e = (y_1^e, \dots, y_m^e)$ i.i.d. with common F_0 , such that

$$E_{F_0} [g(Y, \theta_0)] = 0. \quad (56)$$

For each simulation sample $\mathbf{y}^s(\gamma) = (y_1^s(\gamma), \dots, y_n^s(\gamma))$ with $F(\gamma)$, $\theta(\gamma)$ is the true solution of

$$E_{F(\gamma)} [g(Y, \theta(\gamma))] = 0. \quad (57)$$

Then the true parameter γ_0 is defined by

$$\theta(\gamma_0) = \theta_0 \quad (58)$$

and

$$\Sigma_{g,\gamma}(\theta) = E \left[(g(Y(\gamma), \theta) - E[g(Y(\gamma), \theta)]) (g(Y(\gamma), \theta) - E[g(Y(\gamma), \theta)])' \right]. \quad (59)$$

With the given notation we make the following assumptions.

ASSUMPTION 2.

- (a) $\hat{\theta}_m \xrightarrow{p} \theta_0$, $\theta_0 \in \Theta \subseteq \mathbb{R}^q$ and Θ is compact.
- (b) There is a unique $\gamma_0 \in \Gamma$ such that $\theta(\gamma_0) = \theta_0$ is the solution to the unbiased estimation equation $E_{F(\gamma_0)} [g(Y, \theta_0)] = 0$ and $\Sigma_{g,\gamma_0}(\theta_0)$ has rank q .
- (c) $g(Y, \theta)$ is continuous in a neighborhood of θ_0 .
- (d) $\Sigma_{g,\gamma}(\theta) < \infty$ for all $\gamma \in \Gamma$ and $\theta \in \Theta$.

THEOREM 4. *If Assumption 2 is satisfied then*

$$\hat{\gamma}_{m,n} \xrightarrow{p} \gamma_0$$

as $m, n \rightarrow \infty$.

PROOF. See A.5

5. Conclusion

We have extended the BAEL method introduced by Emerson and Owen (2009), while demonstrating that its limit distribution and its consistency for moment condition models is preserved.

We empirically investigated its capability for simulation-based estimation. The underlying idea of the estimation is to minimize the distance between simulated moments and the empirical moments, using likelihood as a distance measure. This amounts to maximizing the likelihood of observing some empirical moments in simulation outcomes with respect to its configuration. In essence, the estimation approach is a series of hypothesis tests with a fixed hypothesis and varying data sets, generated at different configurations.

In a benchmark problem we attempted to maximize the empirical likelihood of the GBM simulator on average to observe a given empirical moment. The results demonstrate that using eBAEL for simulation based estimation converges to the true parameter and gives good small sample estimates. This is highly attractive for complex simulation models that require compute-intensive simulations. In future work, we will further investigate the property of good small sample estimates in more complex models than GBM. Finally, we also give conditions under which eBAEL provides a consistent simulation based estimator.

References

- Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States* 99(Suppl. 3), 7280–7287.
- Chen, J., A. Variyath, and B. Abraham (2008). Adjusted Empirical Likelihood and its Properties. *Journal of Computational and Graphical Statistics* 17(2), 426–443.
- DiCiccio, T., P. Hall, and J. Romano (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics* 19(2), 1053–1061.
- Emerson, S. and A. Owen (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics* 3, 1161–1192.
- Farmer, J. D. and D. Foley (2009). The economy needs agent-based modelling. *Nature* 460(7256), 685–686.
- Gallant, A. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12(04), 657–681.
- Gourieroux, C. and A. Monfort (1991). Simulation Based Inference in Models with Heterogeneity. *Annals of Economics and Statistics* (20-21), 69–107.
- Hall, P. and B. La Scara (1990). Methods and Algorithms of Empirical Likelihood. *International Statistical Review* 58(2), 109–127.
- Hansen, L. P. and K. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society* 50(5), 1269–1286.
- McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5), 995–1026.
- Newey, W. and R. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.
- Owen, A. (1990). Empirical Likelihood Ratio Confidence Region. *The Annals of Statistics* 18(1), 90–120.
- Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society* 57(5), 1027–1057.
- Qin, J. and J. Lawless (1994). Empirical Likelihood and General Estimating Equations. *The Annals of Statistics* 22(1), 300–325.
- Shalizi, C. R. (2006). Methods and Techniques of Complex Systems Science: An Overview. In T. S. Deisboeck and J. Y. Kresh (Eds.), *Complex Systems Science in Biomedicine*, Topics in Biomedical Engineering International Book Series, Chapter 2, pp. 33–114. Springer US.
- Van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

A. Appendix

Before we come to the main proofs we introduce and discuss some quantities, lemmas and remarks that will be needed later on.

LEMMA 2. *Let Y_i be independent random variables with common distribution and suppose that $E(Y_i^2) < \infty$. Then*

$$\frac{1}{n} \sum_{i=1}^n Y_i^3 = o\left(n^{\frac{1}{2}}\right)$$

with probability 1 as $n \rightarrow \infty$.

PROOF. See [Owen \(1990\)](#), p.98.

REMARK 1. *Let Y_i i.i.d. and suppose a measurable function f with $\Sigma_f = \text{Var}[f(Y)] < \infty$ and $E[f] = 0$. It follows that $f_i = f(Y_i)$ are also i.i.d. and $E[f^2] < \infty$. With Lemma 2 we have*

$$\frac{1}{n} \sum_{i=1}^n f_i^3 = o_p\left(n^{\frac{1}{2}}\right).$$

We introduce the following quantities: a constant s and the function

$$g_{n+2}(\theta) = 2\bar{g}_n(\theta) + sc_u(\theta)u$$

where $\bar{g}_n(\theta) = \sum_{i=1}^n g_i(\theta)$ and $c_u(\theta) = \left(u' \hat{S}(\theta)^{-1} u\right)^{-1/2}$ with

$$\hat{S}(\theta) = \frac{1}{n} \sum_{i=1}^n (g_i(\theta) - \bar{g}_n(\theta))(g_i(\theta) - \bar{g}_n(\theta))'.$$

Moreover, $g_{n+1}(\theta) = -sc_u(\theta)u$ and

$$S(\theta) = \frac{1}{n} \sum_{i=1}^n g_i(\theta) g_i(\theta)',$$

$$\tilde{S}(\theta) = \frac{1}{n+2} \sum_{i=1}^{n+2} g_i(\theta) g_i(\theta)',$$

$$\tilde{\bar{g}}_n(\theta) = \frac{1}{n+2} \sum_{i=1}^{n+2} g_i(\theta).$$

Note that

$$\tilde{\bar{g}}_n(\theta) = \frac{1}{n+2} [n\bar{g}_n(\theta) + 2\bar{g}_n(\theta)] = \bar{g}_n(\theta).$$

LEMMA 3. *For given θ suppose $E[g(Y, \theta)g(Y, \theta)'] < \infty$ and $E[g(Y, \theta)] < \infty$ then*

$$c_u(\theta) = O_p(1).$$

PROOF. Given the assumptions $\Sigma_g(\theta) = E[(g(Y, \theta) - E[g(Y, \theta)])(g(Y, \theta) - E[g(Y, \theta)])'] < \infty$ exists and $\hat{S}(\theta) \xrightarrow{p} \Sigma_g(\theta)$. Since the variance-covariance matrix $\Sigma_g(\theta)$ is positive-semidefinite (p.s.d.) and symmetric, it has positive eigenvalues. Let $\gamma_1(\theta) \geq \dots \geq \gamma_d(\theta)$ be the eigenvalues of $\Sigma_g(\theta)$. As $\hat{S}(\theta) \xrightarrow{p} \Sigma_g(\theta)$, for any unit vector η we have $\gamma_1^{-1}(\theta) + o_p(1) \leq \eta' \hat{S}(\theta)^{-1} \eta \leq \gamma_d^{-1}(\theta) + o_p(1)$. With the latter it follows $c_u(\theta) = O_p(1)$.

LEMMA 4. For given θ suppose $E[g(Y, \theta)g(Y, \theta)'] < \infty$ and $E[g(Y, \theta)] < \infty$ then

$$\tilde{S}(\theta) \xrightarrow{p} S(\theta).$$

PROOF. With $u = \bar{g}_n(\theta) / \|\bar{g}_n(\theta)\|$ it is

$$\begin{aligned} \tilde{S}(\theta) &= \frac{1}{n+2} \left(\sum_{i=1}^n g_i(\theta) g_i(\theta)' + g_{n+1}(\theta) g_{n+1}(\theta)' + g_{n+2}(\theta) g_{n+2}(\theta)' \right) \\ &= \frac{n}{n+2} S(\theta) + \frac{s^2 c_u^2(\theta) + (2 \|\bar{g}_n(\theta)\| + s c_u(\theta))^2}{n+2} uu'. \end{aligned} \quad (60)$$

As $\bar{g}_n \xrightarrow{p} E[g(Y, \theta)]$, \bar{g}_n has order $O_p(1)$. With $s = O(1)$ and the given assumptions, $c_u(\theta) = O_p(1)$, (see Lemma 3) the last term in eq. (60) is of order

$$\begin{aligned} \left[O(1) O_p(1) + (O_p(1) + O(1) O_p(1))^2 \right] O(n^{-1}) &= O_p(1) O(n^{-1}) \\ &= O_p(n^{-1}). \end{aligned}$$

Hence, $\tilde{S}(\theta) - S(\theta) \rightarrow 0$ in probability as $n \rightarrow \infty$.

A.1. Proof of Lemma 1

From definition of $\tilde{W}(\theta)$, w_i^* must satisfy $0 \leq w_i^* \leq 1$. With eq. (22), λ and θ must satisfy

$$1 + \lambda' g_i(\theta) \geq \frac{1}{n+2}$$

for each i . For fixed θ define $D_\theta = \left\{ \lambda : 1 + \lambda' g_i(\theta) \geq \frac{1}{n+2} \right\}$. D_θ is convex and closed, since 0 is in the convex hull of $g_i(\theta)$'s by construction it is also bounded. Moreover,

$$\frac{\partial}{\partial \lambda} \left(\frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i(\theta)}{1 + \lambda' g_i(\theta)} \right) = -\frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i(\theta) g_i(\theta)'}{(1 + \lambda' g_i(\theta))^2}$$

is negative definite for λ in D_θ provided that $\frac{1}{n+2} \sum_{i=1}^{n+2} g_i(\theta) g_i(\theta)'$ is positive definite. By the inverse function theorem $\lambda = \lambda(\theta)$ is a continuous differentiable function. Since $\Sigma_g(\theta_0) < \infty$ with Lemma 5 and WLN we have

$$\tilde{S}(\theta_0) \xrightarrow{p} E[g(Y, \theta_0)g(Y, \theta_0)'] = \Sigma_g(\theta_0)$$

as $n \rightarrow \infty$. As $\tilde{W}(\theta) = -\sum_{i=1}^{n+2} \log(1 + \lambda(\theta)' g_i(\theta))$ with the assumptions and large enough n it is continuous in θ_0 .

A.2. Proof of Theorem 1

The proof in this section is similar to the one in Owen (1990) and Emerson and Owen (2009). However, to the best of our knowledge there has been no proof published to demonstrate the distributional convergence of the BAEI for unbiased estimation equations. Throughout this proof we assume $\theta = \theta_0$ for which we have $E[g(Y, \theta_0)] = E_{F_0}[g] = 0$. For the rest of this section we will write the argument θ_0 only for emphasis, otherwise we will drop the argument for convenience,

i.e. $\bar{g}_n = \sum_{i=1}^n g_i(\theta_0)$. Moreover we define $g^* = \max_{i=1:n} \|g_i\|$, $\tilde{g}^* = \max_{i=1:n+2} \|g_i\|$ and the following magnitudes hold: i) $\dagger g^* = o_p\left(n^{\frac{1}{2}}\right)$, ii) $\ddagger \bar{g}_n = O_p\left(n^{-1/2}\right)$ iii) $g_{n+1} = O_p(1)$, iv) $g_{n+2} = O_p(1)$ and v) $\tilde{g}^* = o_p\left(n^{\frac{1}{2}}\right)$.

Note that by assumption, iii) follows from Lemma 3; that is $c_u = c_u(\theta_0) = O_p(1)$ since Theorem 1 assumes

$$\Sigma_g = E[g(Y, \theta_0) g(Y, \theta_0)'] < \infty.$$

The latter gives $g_{n+1} = O_p(1)$. Moreover using $c_u = O_p(1)$ and $\bar{g}_n = O_p(n^{-1/2})$ with the definition of g_{n+2} , we get

$$\begin{aligned} g_{n+2} &= O_p\left(n^{-1/2}\right) + O_p(1) \\ &= O_p(1). \end{aligned}$$

Finally, as $g_{n+2} = g_{n+1} = O_p(1)$, \tilde{g}^* has the same order as g^* , i.e. $\tilde{g}^* = o_p\left(n^{\frac{1}{2}}\right)$. Before we come to the main proof we need the following lemma and remark.

LEMMA 5. *Let $\theta = \theta_0$ and $E[g(Y, \theta_0) g(Y, \theta_0)'] < \infty$ and $E[g(Y, \theta_0)] < \infty$, then*

$$\tilde{S} - S \xrightarrow{p} 0$$

as $n \rightarrow \infty$.

PROOF. Similarly to Lemma 4, $\tilde{S} = \tilde{S}(\theta_0)$ can be written as

$$\tilde{S} = \frac{n}{n+2} S + \frac{s^2 c_u^2 + (b_n \|\bar{g}_n\| + s c_u)^2}{n+2} u u'.$$

As c_u is of order $O_p(1)$, $s = O(1)$ and from above \bar{g}_n is order of $O_p(n^{-\frac{1}{2}})$, the order of the last term is

$$\begin{aligned} &\left[O(1) O_p(1) + \left(O_p\left(n^{-\frac{1}{2}}\right) + O(1) O_p(1) \right)^2 \right] O(n^{-1}) \\ &= \left[O_p(1) + (o_p(1) + O_p(1))^2 \right] O(n^{-1}) \\ &= [O_p(1) + O_p(1)] O(n^{-1}) \\ &= O(n^{-1}) \end{aligned}$$

Hence $\tilde{S} - S \rightarrow 0$ in probability as $n \rightarrow \infty$.

REMARK 2. *As above for $\theta = \theta_0$ we have $E_{F_0}[g] = 0$ and $\Sigma_g = E[g(Y, \theta_0) g(Y, \theta_0)'] < \infty$, then $S \rightarrow \Sigma_g$ in probability as $n \rightarrow \infty$. Furthermore $\bar{g}_n \rightarrow E_{F_0}[g] = 0$ in probability as $n \rightarrow \infty$, it follows $\frac{1}{n} \sum_{i=1}^n \|g_i\|^2 \rightarrow E_{F_0}[\|g\|^2] < \infty$ in probability as $n \rightarrow \infty$. According to Lemma 5 we have $\tilde{S} \rightarrow \Sigma_g < \infty$ and for $b_n = 2$, it is $\tilde{g}_n = \bar{g}_n \rightarrow E_{F_0}[g] = 0$ in probability as $n \rightarrow \infty$. From the latter two statements it follows $\frac{1}{n+2} \sum_{i=1}^{n+2} \|g_i\|^2 \rightarrow E_{F_0}[\|g\|^2] < \infty$ in probability as $n \rightarrow \infty$.*

\dagger (Chen et al., 2008) p.22.

\ddagger (Qin and Lawless, 1994) p.318.

The remainder of this section proves Theorem 1. The proof is outlined as follows. First we derive that $\|\lambda\| = O_p\left(n^{-\frac{1}{2}}\right)$. Knowing that, we show $\lambda = \tilde{S}^{-1}\tilde{g}_n + o_p\left(n^{-\frac{1}{2}}\right)$, for the sample covariance matrix \tilde{S} . We complete the proof by substituting this expression for λ into the profile (balance adjusted) empirical log likelihood ratio statistic $-2W(\theta_0)$, verifying that some other terms are negligible and using Lemma 5. Accordingly the proof of Theorem 1 is divided into three parts.

Part 1:

PROOF. Without loss of generality let $\sigma_1^2 \leq \dots \leq \sigma_m^2$ be the eigenvalues of $\Sigma_g = E_{F_0}[gg']$ with $\sigma_1^2 = 1$. For $\theta = \theta_0$ using $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ and $\hat{\lambda} = \lambda/\rho$, $\rho = \|\lambda\|$ in eq. (23) it follows

$$\begin{aligned}
0 &= \frac{\hat{\lambda}'}{n+2} \sum_{i=1}^{n+2} \frac{g_i}{1+\hat{\lambda}'g_i} = \frac{\hat{\lambda}'}{n+2} \sum_{i=1}^{n+2} g_i - \frac{\hat{\lambda}'}{n+2} \sum_{i=1}^{n+2} \frac{g_i \hat{\lambda}' g_i}{1+\hat{\lambda}'g_i} \\
&= \hat{\lambda}' \tilde{g}_n - \frac{1}{n+2} \sum_{i=1}^{n+2} \frac{\hat{\lambda}' g_i \rho \hat{\lambda}' g_i}{1+\rho \hat{\lambda}' g_i} \\
&= \hat{\lambda}' \tilde{g}_n - \frac{\rho}{n+2} \sum_{i=1}^{n+2} \frac{\hat{\lambda}' g_i g_i' \hat{\lambda}}{1+\rho \hat{\lambda}' g_i} \\
&\leq \hat{\lambda}' \tilde{g}_n - \frac{\rho}{1+\rho \tilde{g}^*} \hat{\lambda}' \tilde{S} \hat{\lambda} \\
&\leq \hat{\lambda}' \tilde{g}_n - \frac{\rho(1-\varepsilon)}{1+\rho \tilde{g}^*}.
\end{aligned} \tag{61}$$

The last inequality follows from the fact that $\tilde{S} \xrightarrow{P} \Sigma_g$ (,using $S \xrightarrow{P} \Sigma_g$ and Lemma 5). Therefore in probability for some $\varepsilon > 0$ we have

$$\hat{\lambda}' \tilde{S} \hat{\lambda} \geq (1-\varepsilon) \sigma_1^2 = (1-\varepsilon).$$

Using eq. (61) gives

$$\frac{\rho}{(1+\rho \tilde{g}^*)} \leq \frac{\hat{\lambda}' \tilde{g}_n}{1-\varepsilon}. \tag{62}$$

Since $\tilde{g}_n = \bar{g}_n$ and $\frac{\hat{\lambda}' \tilde{g}_n}{1-\varepsilon}$ is of order $O_p\left(n^{-\frac{1}{2}}\right)$ with eq. (62) it follows

$$\rho = \|\lambda\| = O_p\left(n^{-\frac{1}{2}}\right). \tag{63}$$

Part 2:

PROOF. First define $\vartheta_i = \lambda' g_i$. Having established an order bound for $\|\lambda\|$ and with $\tilde{g}^* = o_p\left(n^{\frac{1}{2}}\right)$ it is

$$\max_{i=1:n+2} |\vartheta_i| = O_p\left(n^{-\frac{1}{2}}\right) o_p\left(n^{\frac{1}{2}}\right) = o_p(1). \tag{64}$$

Using $\frac{1}{1+x} = 1 - x - \frac{x^2}{1+x}$ in eq. (23), we get

$$\begin{aligned}
0 &= \frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i}{1+\lambda' g_i} = \frac{1}{n+2} \sum_{i=1}^{n+2} g_i \left(1 - \lambda' g_i + \frac{(\lambda' g_i)^2}{1+\lambda' g_i}\right) \\
&= \tilde{g}_n - \tilde{S} \lambda + \frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i (\lambda' g_i)^2}{1+\lambda' g_i}.
\end{aligned} \tag{65}$$

The last term is bounded above by norm

$$\begin{aligned} \frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i (\lambda' g_i)^2}{1 + \lambda' g_i} &\leq \max_{i=1:n+2} \|g_i\| \frac{1}{n+2} \sum_{i=1}^{n+2} \|\lambda\|^2 \|g_i\|^2 \left| 1 + \lambda' g_i \right|^{-1} \\ &= \tilde{g}^* \|\lambda\|^2 \frac{1}{n+2} \sum_{i=1}^{n+2} \|g_i\|^2 \left| 1 + \lambda' g_i \right|^{-1}. \end{aligned} \quad (66)$$

With the given order of \tilde{g}^* and λ , Remark 2 and eq. (64), the order of eq. (66) becomes

$$o_p \left(n^{\frac{1}{2}} \right) \left(O_p \left(n^{-\frac{1}{2}} \right) \right)^2 O_p(1) O_p(1) = o_p \left(n^{-\frac{1}{2}} \right).$$

Using the latter in eq. (65) gives

$$\lambda = \tilde{S}^{-1} \tilde{g}_n + o_p \left(n^{-\frac{1}{2}} \right). \quad (67)$$

Part 3:

PROOF. By eq. (64) we may expand

$$\log(1 + \vartheta_i) = \vartheta_i - \frac{1}{2} \vartheta_i^2 + \eta_i, \quad (68)$$

where for some finite $B > 0$,

$$P \left(|\eta_i| \leq B |\vartheta_i|^3, 1 \leq i \leq n+2 \right) \rightarrow 1 \quad (69)$$

as $n \rightarrow \infty$. Substituting (68) in eq. (24) we get

$$-2\tilde{W}(\theta_0) = 2 \sum_{i=1}^{n+2} \log(1 + \vartheta_i) = 2 \sum_{i=1}^{n+2} \vartheta_i - \sum_{i=1}^{n+2} \vartheta_i^2 + 2 \sum_{i=1}^{n+2} \eta_i.$$

Remark§ 1 and eq. (69) give an order bound for the last term

$$\begin{aligned} 2 \left| \sum_{i=1}^{n+2} \eta_i \right| &\leq 2B \|\lambda\|^3 \sum_{i=1}^{n+2} \|g_i\|^3 \\ &= 2B \|\lambda\|^3 \left[\sum_{i=1}^n \|g_i\|^3 + \|g_{n+1}\|^3 + \|g_{n+2}\|^3 \right] \\ &= 2B O_p \left(n^{-\frac{1}{2}} \right)^3 \left[o_p \left(n^{\frac{3}{2}} \right) + O_p(1) + O_p(1) \right] \\ &= 2B O_p \left(n^{-\frac{3}{2}} \right) \left[o_p \left(n^{\frac{3}{2}} \right) \right] = o_p(1). \end{aligned} \quad (70)$$

Let us rewrite eq. (70) by

$$\lambda = \tilde{S}^{-1} \tilde{g}_n + \beta,$$

§Under the mild condition of g being a measurable function, it follows with Lemma 2 that $\sum_{i=1}^n \|g_i\|^3 = o \left(n^{\frac{3}{2}} \right)$ as $\frac{1}{n} \sum_{i=1}^n \|g_i\|^3 = o \left(n^{\frac{1}{2}} \right)$.

with $\|\beta\| = o_p\left(n^{-\frac{1}{2}}\right)$. Using the latter and re-substituting $\vartheta_i = \lambda' g_i$ in eq. (24) gives

$$\begin{aligned}
-2\tilde{W}(\theta_0) &= 2 \sum_{i=1}^{n+2} \lambda' g_i - \sum_{i=1}^{n+2} (\lambda' g_i)^2 + o_p(1) \\
&= 2(n+2) \lambda' \tilde{g}_n - (n+2) \lambda' \tilde{S} \lambda + o_p(1) \\
&= 2(n+2) \left(\tilde{S}^{-1} \tilde{g}_n + \beta \right)' \tilde{g}_n - (n+2) \left(\tilde{S}^{-1} \tilde{g}_n + \beta \right)' \tilde{S} \left(\tilde{S}^{-1} \tilde{g}_n + \beta \right) + o_p(1) \\
&= 2(n+2) \left[\tilde{g}_n' \tilde{S}^{-1} \tilde{g}_n + \beta' \tilde{g}_n \right] - (n+2) \left[\tilde{g}_n' \tilde{S}^{-1} \tilde{g}_n + 2\beta' \tilde{g}_n + \beta' \tilde{S} \beta \right] + o_p(1) \\
&= (n+2) \left[\tilde{g}_n' \tilde{S}^{-1} \tilde{g}_n \right] + (n+2) \beta' \tilde{S} \beta + o_p(1) \\
&= (n+2) \left[\tilde{g}_n' \tilde{S}^{-1} \tilde{g}_n \right] + o_p(1). \tag{71}
\end{aligned}$$

As $\tilde{S} = O_p(1)$ (using Lemma 5 and $S \xrightarrow{P} \Sigma_g$), the last equality holds because $\tilde{g}_n = \bar{g}_n'$ and

$$(n+2) \beta' \tilde{S} \beta = O(n) o_p\left(n^{-1/2}\right) O_p(1) o_p\left(n^{-1/2}\right) = o_p(1).$$

Moreover, as $n\bar{g}_n' S^{-1} \bar{g}_n$ converges to a χ^2 distribution with q degrees of freedom, $\tilde{S} \xrightarrow{P} S$ and $\frac{n}{n+2} \rightarrow 1$, it follows $-2\tilde{W}(\theta_0) \rightarrow \chi_q^2$ in probability as $n \rightarrow \infty$.

A.3. Proof of Theorem 2

PROOF. Suppose $\theta \neq \theta_0$. As before we drop the argument θ , e.g. $\bar{g}_n = \frac{1}{n} \sum_{i=1}^n g(y_i, \theta) = \frac{1}{n} \sum_{i=1}^n g_i$, $g_{n+1} = -sc_u(\theta)u$ and $g_{n+2} = 2\bar{g}_n(\theta) + sc_u(\theta)u$. Note, due to the law of large numbers, $\|\bar{g}_n' \bar{g}_n\| \rightarrow \delta^2$ and $\bar{g}_n \rightarrow \mu(\theta) := E[g(Y, \theta)]$ in probability as $n \rightarrow \infty$. By assumption $\Sigma_g(\theta) < \infty$, with Lemma 3 we have $c_u = O_p(1)$. As $E[g(Y, \theta)g(Y, \theta)'] = \Sigma_g(\theta) + \mu(\theta)\mu(\theta)' < \infty$ and $S \xrightarrow{P} E[g(Y, \theta)g(Y, \theta)']$ with Lemma 4 ($\tilde{S} \xrightarrow{P} S$) it follows $\tilde{S} = O_p(1)$.

Now, for $i = 1, \dots, n$ the terms $g_i - \bar{g}_n$ have expected value zero

$$E[g_i - \bar{g}_n] = 0$$

and satisfying all moment conditions such that with Lemma 3 in (Owen, 1990, p. 98) it follows that

$$\max_{i=1, \dots, n} \{\|g_i - \bar{g}_n\|\} = o_p\left(n^{1/2}\right). \tag{72}$$

Let $\tilde{\lambda} = n^{-2/3} \bar{g}_n M$ for a positive constant $M > 0$. For $i = 1, \dots, n$

$$\tilde{\lambda}' g_i = \tilde{\lambda}' (g_i - \bar{g}_n) + \tilde{\lambda}' \bar{g}_n. \tag{73}$$

From the above \bar{g}_n is of order $O_p(1)$ therefore the maximum of the first term on the right hand side in eq. (73) is with eq. (72) of order $o_p\left(n^{-2/3} n^{1/2}\right) = o_p(1)$. The last term in eq. (73) has the order $n^{-2/3} O_p(1) = o_p(1)$ hence

$$\max_{i=1, \dots, n} \left\{ \|\tilde{\lambda}' g_i\| \right\} = o_p(1). \tag{74}$$

Since s and u are of $O(1)$ and $c_u = O_p(1)$ it follows that $g_{n+1} = O_p(1)$ and $g_{n+2} = O_p(1)$. Hence $\tilde{\lambda}' g_{n+1} = o_p(1)$ and $\tilde{\lambda}' g_{n+2} = o_p(1)$ therefore

$$\max_{i=1: n+2} \left\{ \|\tilde{\lambda}' g_i\| \right\} = o_p(1). \tag{75}$$

With eq. (75) for $i = 1, \dots, n+2$ we have $1 + \tilde{\lambda}'g_i > 0$ with probability going to 1. Hence using the Taylor expansion:

$$\log(1+x) = x - \frac{x^2}{2(1+\xi)^2} \quad (76)$$

for some ξ between 0 and x and the duality of the maximization problem it is

$$\begin{aligned} \tilde{W}(\theta) &= -\sup_{\lambda} \left\{ \sum_{i=1}^{n+2} \log(1 + \lambda'g_i) \right\} \\ &\leq -\sum_{i=1}^{n+2} \log(1 + \tilde{\lambda}'g_i) \\ &= -\left[\sum_{i=1}^{n+2} \tilde{\lambda}'g_i - \frac{1}{2} \sum_{i=1}^{n+2} \frac{(\tilde{\lambda}'g_i)^2}{(1+\xi_i)^2} \right]. \end{aligned} \quad (77)$$

Note, from eq. (75) all ξ_i are within $o_p(1)$ neighborhood of 0 uniformly. Therefore the second term in the last line of eq. (77) is no larger than

$$\sum_{i=1}^{n+2} (\tilde{\lambda}'g_i)^2 = (n+2) \tilde{\lambda}'\tilde{S}\tilde{\lambda} = O(n) O_p(n^{-2/3}) O_p(1) O_p(n^{-2/3}) = o_p(1).$$

The first term is

$$\sum_{i=1}^{n+2} \tilde{\lambda}'g_i = \tilde{\lambda}'n\bar{g}_n + 2\tilde{\lambda}'\bar{g}_n = n^{1/3}\delta^2M + o_p(1).$$

Therefore eq. (77) gives

$$\tilde{W}(\theta) \leq -n^{1/3}\delta^2M + o_p(1). \quad (78)$$

Since M can be arbitrarily large, we have $-2n^{-1/3}\tilde{W}(\theta) \rightarrow \infty$ for any $\theta \neq \theta_0$.

A.4. Proof of Theorem 3

LEMMA 6. Suppose $E \left[\sup_{\theta \in \Theta} \|g(Y, \theta)\|^\alpha \right] < \infty$ for some $\alpha > 2$ then

$$\Sigma_g(\theta) < \infty$$

for all $\theta \in \Theta$.

PROOF. From $E \left[\sup_{\theta \in \Theta} \|g(Y, \theta)\|^\alpha \right] < \infty$ for some $\alpha > 2$ it follows $E \left[\|g(Y, \theta)\|^2 \right] < \infty$ for all $\theta \in \Theta$. As $g(Y, \theta) = (g_1(Y, \theta), \dots, g_l(Y, \theta))' = (g_1, \dots, g_l)'$ the latter implies $E[g_i^2] < \infty$, hence $g_i \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ for all $\theta \in \Theta$ and $i = 1, \dots, l$. As a consequence

$$E[g_i g_j] = \int_{\Omega} g_i g_j dP \leq \int_{\Omega} \|g_i g_j\| dP \leq \left(\int_{\Omega} \|g_i\|^2 dP \right)^{1/2} \left(\int_{\Omega} \|g_j\|^2 dP \right)^{1/2} < \infty \quad (79)$$

for all $i, j = 1, \dots, l$ and $\theta \in \Theta$, where we used the Hoelder inequality and the fact that $g_i \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ for all $\theta \in \Theta$ and $i = 1, \dots, l$. Note the variance covariance matrix can be written as

$$\Sigma_g(\theta) = E[gg'] - E[g]E[g'].$$

By assumption it follows $E[g] < \infty$ hence with eq. (79) we have $\Sigma_g(\theta) < \infty$ for all $\theta \in \Theta$.

REMARK 3. As a consequence of Lemma 6, Assumption 1 satisfy the conditions of Lemma 3, hence $c_u(\theta) = O_p(1)$ for all $\theta \in \Theta$. Moreover, $\bar{g}_n(\theta) = O_p(1)$ for all $\theta \in \Theta$. Altogether this results in $\|g_{n+1}(\theta)\| = \|g_{n+2}(\theta)\| = O_p(1)$ for all $\theta \in \Theta$.

The proof of Theorem 3 is based on Newey and Smith (2004) and is divided into four parts (three Lemmas and the main proof).

LEMMA 7. If Assumption 1 is satisfied, then for any ζ with $1/\alpha < \zeta < 1/2$ and $\Lambda_n = \{\lambda : \|\lambda\| \leq n^{-\zeta}\}$, we have $\sup_{\theta \in \Theta, \lambda \in \Lambda_n, i=1, \dots, n+2} |\lambda' g_i(\theta)| \xrightarrow{P} 0$ and with probability approaching (w.p.a.) 1, $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ for all $\theta \in \Theta$.

PROOF.

$$\begin{aligned} \sup_{\theta \in \Theta, \lambda \in \Lambda_n, i=1, \dots, n+2} |\lambda' g_i(\theta)| &\leq \sup_{\lambda \in \Lambda_n} \|\lambda\| \max_{i=1, \dots, n+2} \sup_{\theta \in \Theta} \|g_i(\theta)\| \\ &= \sup_{\lambda \in \Lambda_n} \|\lambda\| \left(\max_{i=1, \dots, n+2} \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha \right)^{1/\alpha} \\ &\leq \sup_{\lambda \in \Lambda_n} \|\lambda\| \left(\sum_{i=1}^{n+2} \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha \right)^{1/\alpha} \\ &= n^{1/\alpha} \sup_{\lambda \in \Lambda_n} \|\lambda\| \left(\frac{1}{n} \sum_{i=1}^{n+2} \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha \right)^{1/\alpha} \\ &= n^{1/\alpha} \sup_{\lambda \in \Lambda_n} \|\lambda\| \left(\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha + \frac{1}{n} \sum_{i=n+1}^{n+2} \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha \right)^{1/\alpha} \\ &= n^{1/\alpha} O(n^{-\zeta}) (O_p(1) + O_p(n^{-1})) \\ &= o_p(1) \end{aligned}$$

The second to last line holds due to Remark 3 and the assumption $E \left[\sup_{\theta \in \Theta} \|g(y, \theta)\|^\alpha \right] < \infty$ for some $\alpha > 2$ that gives $\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha \xrightarrow{P} E \left[\sup_{\theta \in \Theta} \|g(y, \theta)\|^\alpha \right] < \infty$, i.e.

$$\frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \Theta} \|g_i(\theta)\|^\alpha = O_p(1).$$

Overall it follows w.p.a. $1 \lambda' g_i(\theta) \in \mathfrak{F}$ for all $\theta \in \Theta$ and $\|\lambda\| \leq n^{-\zeta}$.

LEMMA 8. If Assumption 1 is satisfied, $\bar{\theta} \in \Theta$, with $\bar{\theta} \xrightarrow{P} \theta_0$ and $\tilde{g}_n(\bar{\theta}) = O_p(n^{-1/2})$, then $\hat{\lambda} = \operatorname{argmax}_{\lambda \in \hat{\Lambda}_n(\bar{\theta})} \hat{P}(\bar{\theta}, \lambda) = \operatorname{argmax}_{\lambda \in \hat{\Lambda}_n(\bar{\theta})} \sum_{i=1}^{n+2} \rho(\lambda' g_i(\bar{\theta})) / (n+2)$ exists with w.p.a. 1, $\hat{\lambda} = O_p(n^{-1/2})$ and $\sup_{\lambda \in \hat{\Lambda}_n(\bar{\theta})} \hat{P}(\bar{\theta}, \lambda) \leq \rho_0 + O_p(n^{-1})$.

PROOF. Since Assumption 1 satisfy the conditions of Lemma 4 we have $\tilde{S}(\bar{\theta}) \xrightarrow{P} S(\bar{\theta})$. By assumptions and the UWL (Uniform Weak Law of Large Numbers) we have $S(\bar{\theta}) \xrightarrow{P} \sum_g(\theta_0)$, hence $\tilde{S}(\bar{\theta}) \xrightarrow{P} \sum_g(\theta_0)$. By $\sum_g(\theta_0) < \infty$ the smallest eigenvalue $\tilde{S}(\bar{\theta})$ is bounded away from 0 w.p.a. 1. Since $\rho(v)$ is twice continuously differentiable in the neighborhood of 0 with Lemma 7 it follows $\hat{P}(\bar{\theta}, \lambda)$ is twice continuously differentiable on Λ_n with w.p.a. 1. Hence, $\check{\lambda} = \operatorname{argmax}_{\lambda \in \Lambda_n} \hat{P}(\bar{\theta}, \lambda)$ exists with w.p.a. 1. Furthermore, for $\bar{g}_i = g_i(\bar{\theta})$ and any $\check{\lambda}$ on the line joining $\check{\lambda}$ and 0 it follows from Lemma 7 and $\rho_2 = -1$ that $\max_{1 \leq i \leq n+2} \rho_2(\check{\lambda} \bar{g}_i) < -1/2$ with w.p.a. 1. Then using the Taylor Expansion of $\hat{P}(\bar{\theta}, \lambda)$ around $\lambda = 0$ and $\check{\lambda}$ on the line joining $\check{\lambda}$ and 0 we get

$$\begin{aligned} \rho_0 = \hat{P}(\bar{\theta}, 0) \leq \hat{P}(\bar{\theta}, \check{\lambda}) &= \rho_0 - \check{\lambda}' \tilde{g}_n(\bar{\theta}) + \frac{1}{2} \check{\lambda}' \left[\sum_{i=1}^{n+2} \rho_2(\check{\lambda}' \bar{g}_i) \bar{g}_i \bar{g}_i' / (n+2) \right] \check{\lambda} \\ &\leq \rho_0 - \check{\lambda}' \tilde{g}_n(\bar{\theta}) - \frac{1}{4} \check{\lambda}' \tilde{S}(\bar{\theta}) \check{\lambda} \\ &\leq \rho_0 + \|\check{\lambda}\| \|\tilde{g}_n(\bar{\theta})\| - C_1 \|\check{\lambda}\|^2, \end{aligned} \quad (80)$$

where C_1 a positive constant. Subtracting $\rho_0 - C_1 \|\check{\lambda}\|^2$ from both sides and dividing $\|\check{\lambda}\|^2$ we get $C_1 \|\check{\lambda}\| \leq \|\tilde{g}_n\|$ w.p.a. 1. By assumption we have $\tilde{g}_n(\bar{\theta}) = O_p(n^{-1/2})$, therefore $\|\check{\lambda}\| = O_p(n^{-1/2}) = o_p(n^{-\zeta})$. From the latter it follows that $\check{\lambda} \in \operatorname{int}(\Lambda_n)$ w.p.a.1 and with Lemma 7 $\check{\lambda} \in \hat{\Lambda}_n(\bar{\theta})$ w.p.a. 1. By concavity of $\hat{P}(\bar{\theta}, \lambda)$ and convexity of $\hat{\Lambda}_n(\bar{\theta})$ it follows $\hat{P}(\bar{\theta}, \check{\lambda}) = \sup_{\lambda \in \hat{\Lambda}_n(\bar{\theta})} \hat{P}(\bar{\theta}, \lambda)$ and therefore $\check{\lambda} = \hat{\lambda}$. Using $\tilde{g}_n(\bar{\theta}) = O_p(n^{-1/2})$, $\|\check{\lambda}\| = O_p(n^{-1/2})$ and in eq. (80) we get

$$\hat{P}(\bar{\theta}, \check{\lambda}) \leq \rho_0 + \|\check{\lambda}\| \|\tilde{g}_n(\bar{\theta})\| - C_1 \|\check{\lambda}\|^2 = \rho_0 + O_p(n^{-1}).$$

LEMMA 9. *If Assumption 1 holds, then*

$$\|\tilde{g}_n(\hat{\theta})\| = O_p(n^{-1/2}).$$

PROOF. Let $\hat{g}_i = g_i(\hat{\theta})$, $\hat{g} = \tilde{g}_n(\hat{\theta})$ and for ζ in Lemma 7, $\check{\lambda} = -n^{-\zeta} \hat{g} / \|\hat{g}\|$. With Lemma 7 it follows $\max_{1 \leq i \leq n+2} |\check{\lambda}' \hat{g}_i| \xrightarrow{P} 0$ and $\check{\lambda} \in \hat{\Lambda}_n(\hat{\theta})$ w.p.a. 1. Then for any $\check{\lambda}$ on the line joining $\check{\lambda}$ and 0 w.p.a. 1 we have $\rho_2(\check{\lambda}' \hat{g}_i) \geq -C_2$ for all $i = 1, \dots, n+2$, where C_2 is a positive constant. Given Assumption 1, Lemma (4) gives $\frac{1}{n+2} \sum_{i=1}^{n+2} \hat{g}_i \hat{g}_i' \xrightarrow{P} \frac{1}{n} \sum_i \hat{g}_i \hat{g}_i'$ and by CS (Cauchy-Schwarz inequality) and UWL it is $\frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}_i' \leq \left(\frac{1}{n} \sum_{i=1}^n (\sup_{\theta \in \Theta} \|g_i(\theta)\|)^2 \right) I \xrightarrow{P} C_3 I$, where C_3 is a positive constant. From the latter it follows that the largest eigenvalue of $\frac{1}{n+2} \sum_i \hat{g}_i \hat{g}_i'$ is bounded above w.p.a. 1. Using Taylor Expansion as before

$$\begin{aligned} \hat{P}(\hat{\theta}, \check{\lambda}) &= \rho_0 - \check{\lambda}' \hat{g} + \frac{1}{2} \check{\lambda}' \left[\sum_{i=1}^{n+2} \rho_2(\check{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i' / (n+2) \right] \check{\lambda} \\ &\geq \rho_0 + n^{-\zeta} \|\hat{g}\| - C_2 \frac{1}{2} \check{\lambda}' \left[\sum_{i=1}^{n+2} \hat{g}_i \hat{g}_i' / (n+2) \right] \check{\lambda} \\ &\geq \rho_0 + n^{-\zeta} \|\hat{g}\| - C n^{-2\zeta} \end{aligned} \quad (81)$$

w.p.a. 1, where $C = C_2 C_3$. By the Lindeberg-Levy central limit theorem the hypothesis of Lemma 8 are satisfied by $\hat{\theta} = \theta_0$. As $\hat{\theta}$ and $\hat{\lambda}$ being saddle point solutions, eq. (81) and Lemma 8 gives:

$$\rho_0 + n^{-\zeta} \|\hat{g}\| - Cn^{-2\zeta} \leq \hat{P}(\hat{\theta}, \hat{\lambda}) \leq \hat{P}(\hat{\theta}, \hat{\lambda}) \leq \sup_{\lambda \in \hat{\Lambda}_n(\theta_0)} \hat{P}(\theta_0, \lambda) \leq \rho_0 + O_p(n^{-1}). \quad (82)$$

Solving the latter for $\|\hat{g}\|$ gives

$$\|\hat{g}\| \leq O_p(n^{\zeta-1}) + Cn^{-\zeta} = O_p(n^{-\zeta}). \quad (83)$$

The last equality holds because by assumption $\zeta < 1/2$, thus $\zeta - 1 < -1/2 < -\zeta$. Now consider $\varepsilon_n \rightarrow 0$ and let $\hat{\lambda} = -\varepsilon_n \hat{g}$, with eq. (83) $\hat{\lambda} = o_p(n^{-\zeta})$, so that $\hat{\lambda} \in \Lambda_n$ w.p.a. 1. Then as in eq. (82)

$$\rho_0 - \hat{\lambda}' \hat{g} - C \|\hat{\lambda}\|^2 = \rho_0 + \varepsilon_n \|\hat{g}\|^2 - C\varepsilon_n^2 \|\hat{g}\|^2 = \rho_0 + (1 - C\varepsilon_n) \varepsilon_n \|\hat{g}\|^2 \leq \rho_0 + O_p(n^{-1}).$$

Since for large enough n , $1 - C_1 \varepsilon_n$ is bounded away from 0 w.p.a. 1 and it follows from the latter equation, $\varepsilon_n \|\hat{g}\|^2 = O_p(n^{-1})$. The final conclusion follows by standard result from probability theory, that if $\varepsilon_n Y_n = O_p(n^{-1})$ for all $\varepsilon_n \rightarrow 0$ then $Y_n = O_p(n^{-1})$.

Provided with the given Lemma 7-9 the following proofs Theorem 3.

PROOF. First note, $\tilde{g}_n(\theta) = \bar{g}_n(\theta)$ then

$$\|\tilde{g}_n(\hat{\theta}) - E[g(Y, \hat{\theta})]\| = \|\bar{g}_n(\hat{\theta}) - E[g(Y, \hat{\theta})]\| \leq \sup_{\theta \in \Theta} \|\bar{g}_n(\theta) - E[g(Y, \theta)]\| \xrightarrow{p} 0,$$

where the latter follows from the assumptions and the UWL. As Lemma 9 gives $\tilde{g}_n(\hat{\theta}) \xrightarrow{p} 0$ it follows from above $E[g(Y, \hat{\theta})] \xrightarrow{p} 0$. By assumption $E[g(Y, \theta)] = 0$ has a unique solution at θ_0 , hence $\|E[g(Y, \theta)]\|$ must be bounded away from 0 outside any neighborhood of θ_0 . Therefore $\hat{\theta}$ must be inside any neighborhood of θ_0 w.p.a. 1, i.e. $\hat{\theta} \xrightarrow{p} \theta_0$. With Lemma 9 ($\|\tilde{g}_n(\hat{\theta})\| = O_p(n^{-1/2})$) and $\bar{\theta} = \hat{\theta}$ the hypotheses in Lemma 8 are satisfied, hence $\hat{\lambda} = \operatorname{argmax}_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \hat{P}(\hat{\theta}, \lambda) = \operatorname{argmax}_{\lambda \in \hat{\Lambda}_n(\hat{\theta})} \sum_{i=1}^{n+2} \rho(\lambda' g_i(\hat{\theta})) / (n+2)$ exists with w.p.a. 1, $\hat{\lambda} = O_p(n^{-1/2})$.

A.5. Proof of Theorem 4

Before we come to the main proof we need to establish various preliminary results. The consistency proof of $\hat{\gamma}_{m,n}$ relies essentially on i) $-\tilde{W}_\gamma(\hat{\theta}_m)$ diverges in probability for $\gamma \neq \gamma_0$ and ii) $-2\tilde{W}_{\gamma_0}(\hat{\theta}_m)$ is bounded in probability. The second property is a consequence of $-2\tilde{W}_{\gamma_0}(\theta_0)$ converging in distribution to a χ_q^2 random variable and $\tilde{W}_{\gamma_0}(\hat{\theta}_m)$ converging in probability to $\tilde{W}_{\gamma_0}(\theta_0)$. The latter will be demonstrated in Lemma 12. The first property will be demonstrated in Lemma 11, using the fact that $-n^{-1/3} \tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability for certain $\theta \in \Theta$. In particular let $\theta(\gamma)$ denote a solution of $E_{F(\gamma)}[g(Y, \theta(\gamma))] = 0$. For the case $\theta(\gamma)$ is an element

¶Note, $\tilde{g}_n(\theta_0) = \bar{g}_n(\theta_0)$.

of Θ with Remark 1 it follows $-n^{-1/3}\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability as $n \rightarrow \infty$ for all $\theta \in \Theta$ with $\theta \neq \theta(\gamma)$. If $\theta(\gamma) \notin \Theta$, we get with Remark 1, $-n^{-1/3}\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability for all $\theta \in \Theta$ as $n \rightarrow \infty$. For generalization consider the case where for $\gamma \neq \gamma_0$ there exist various $\theta^i(\gamma) \in \mathbb{R}^q$ with $i = 1, \dots, p$ such that $E_{F(\gamma)}[g(Y, \theta^i(\gamma))] = 0$. Define the set $\Theta_\gamma^C := \Theta \setminus \Theta^*$ with $\Theta^* := \{\theta^i(\gamma) \mid \theta^i(\gamma) \in \Theta\}$, where Θ_γ^C contains all $\theta \in \Theta \wedge \theta \neq \theta^i(\gamma)$ for $i = 1, \dots, p$. As before with Remark 1 we get $-n^{-1/3}\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability as $n \rightarrow \infty$ for all $\theta \in \Theta_\gamma^C$. Now, as Lemma 10 will demonstrate $\hat{\theta}_m$ is an element of Θ_γ^C w.p.a 1 as $m \rightarrow \infty$, it follows in Lemma 11 that $-\tilde{W}_\gamma(\hat{\theta}_m) \rightarrow \infty$ in probability as $m, n \rightarrow \infty$.

LEMMA 10. Suppose for some $\gamma \neq \gamma_0$ there are $\theta^i(\gamma)$ with $i = 1, \dots, p$ such that $E_{F(\gamma)}[g(Y, \theta^i(\gamma))] = 0$. Suppose $\hat{\theta}_m \rightarrow \theta_0$ in probability as $m \rightarrow \infty$ with $\theta_0 \in \Theta$, then

$$P(\hat{\theta}_m \in \Theta_\gamma^C) \rightarrow 1$$

as $m \rightarrow \infty$, where Θ_γ^C is the set that contains all $\theta \in \Theta$ with $\theta \neq \theta^i(\gamma)$ for $i = 1, \dots, p$.

PROOF. By assumption we have $\hat{\theta} \xrightarrow{P} \theta_0$ with $\theta_0 \in \Theta$, i.e.

$$P(\|\hat{\theta}_m - \theta_0\| \leq \varepsilon) \rightarrow 1 \quad (84)$$

for all $\varepsilon > 0$ as $m \rightarrow \infty$. Therefore, w.p.a. 1, $\hat{\theta}_m \in \Theta$. Let $\Delta = \min_{\theta^i(\gamma) \in \Theta^*} \|\theta^i(\gamma) - \theta_0\|$, as $\gamma \neq \gamma_0$ it is $\theta^i(\gamma) \neq \theta_0$ for $i = 1, \dots, p$, therefore $\Delta > 0$. Now choose any $\varepsilon^* < \Delta$, then $\delta^* := \Delta - \varepsilon^* > 0$. Because $\Delta = \min_{\theta^i(\gamma) \in \Theta^*} \|\theta^i(\gamma) - \theta_0\|$, δ^* is a lower bound of the distance between $\hat{\theta}_m$ and any $\theta^i(\gamma) \in \Theta^*$. With eq. (84) it follows

$$P(\|\hat{\theta}_m - \theta^i(\gamma)\| \geq \delta^*) \rightarrow 1$$

for all $\theta^i(\gamma) \in \Theta^*$ as $m \rightarrow \infty$. The latter means w.p.a. 1, $\hat{\theta}_m$ has at least a $\delta^* > 0$ distance to each $\theta^i(\gamma) \in \Theta^*$. And as $\hat{\theta}_m \in \Theta$ w.p.a. 1, overall we have

$$P(\hat{\theta}_m \in \Theta_\gamma^C) \rightarrow 1$$

as $m \rightarrow \infty$.

LEMMA 11. Suppose $\Sigma_{g,\gamma}(\theta) < \infty$ for all $\theta \in \Theta$ and $\gamma \in \Gamma$ and $\hat{\theta}_m \rightarrow \theta_0$ in probability for $\theta_0 \in \Theta$ as $m \rightarrow \infty$. Suppose $\gamma \neq \gamma_0$, then for any constant M

$$P(-\tilde{W}_\gamma(\hat{\theta}_m) > M) \rightarrow 1$$

as $m, n \rightarrow \infty$.

PROOF. For some $\gamma \neq \gamma_0$ suppose there are $\theta^i(\gamma)$ with $i = 1, \dots, p$ such that $E_{F(\gamma)}[g(Y, \theta^i(\gamma))] = 0$. Then Remark 1 implies $-n^{-1/3}\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability $n \rightarrow \infty$ for all $\theta \in \Theta_\gamma^C$, as Θ_γ^C

||The statement can be derived from the following implications. Let $\ddot{\theta} \in \Theta$, then $x \in \Theta \wedge x \neq \ddot{\theta}$ is equivalent to $x \in \Theta \setminus \{\ddot{\theta}\}$. If $\ddot{\theta} \notin \Theta$, then $x \in \Theta \wedge x \neq \ddot{\theta}$ is equivalent to $x \in \Theta$.

contains all $\theta \in \Theta \wedge \theta \neq \theta^i(\gamma)$ for $i = 1, \dots, p$. As $n^{-1/3} \leq 1$ for $n \in \mathbb{N}$ we have $-\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability for all $\theta \in \Theta_\gamma^C$ as $n \rightarrow \infty$. Hence, for any constant M

$$P\left(-\tilde{W}_\gamma(\theta) > M | \theta \in \Theta_\gamma^C\right) \rightarrow 1$$

as $n \rightarrow \infty$. The latter also holds for any $\hat{\theta}_m$, hence

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M | \hat{\theta}_m \in \Theta_\gamma^C\right) \rightarrow 1 \quad (85)$$

as $n \rightarrow \infty$ and $m \in \mathbb{N}$. Note**,

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M\right) \geq P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M | \hat{\theta}_m \in \Theta_\gamma^C\right) P\left(\hat{\theta}_m \in \Theta_\gamma^C\right). \quad (86)$$

As we assume $\hat{\theta}_m \rightarrow \theta_0$ in probability as $m \rightarrow \infty$ from Lemma 10 we know $P\left(\hat{\theta}_m \in \Theta_\gamma^C\right) \rightarrow 1$ as $m \rightarrow \infty$. Using the latter and eq. (85) in eq. (86) it follows

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M\right) \rightarrow 1$$

as $m, n \rightarrow \infty$. Now suppose for $\gamma \neq \gamma_0$ there is no solution $\theta(\gamma)$ such that $E_{F(\gamma)}[g(Y, \theta(\gamma))] = 0$. Then with Remark 1 we have $-n^{-1/3}\tilde{W}_\gamma(\theta) \rightarrow \infty$ in probability for all $\theta \in \Theta$. Similarly as before we get for any $\hat{\theta}_m$

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M | \hat{\theta}_m \in \Theta\right) \rightarrow 1$$

as $n \rightarrow \infty$ and $m \in \mathbb{N}$. Since $P\left(\hat{\theta}_m \in \Theta\right) \rightarrow 1$ follows from $\hat{\theta}_m \rightarrow \theta_0$ in probability as $m \rightarrow \infty$ and

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M\right) \geq P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M | \hat{\theta}_m \in \Theta\right) P\left(\hat{\theta}_m \in \Theta\right) \quad (87)$$

we have again

$$P\left(-\tilde{W}_\gamma(\hat{\theta}_m) > M\right) \rightarrow 1$$

as $m, n \rightarrow \infty$.

LEMMA 12. Suppose the conditions of Lemma 1 are satisfied and $\hat{\theta}_m \rightarrow \theta_0$ in probability as $m \rightarrow \infty$, then

$$\tilde{W}_{\gamma_0}(\hat{\theta}_m) \rightarrow \tilde{W}_{\gamma_0}(\theta_0)$$

in probability as $n, m \rightarrow \infty$.

PROOF. This proof is similar to the CMT (Continuous Mapping Theorem) proof in Van der Vaart (1998) (see p. 8). For the following we will write $\tilde{W}_{\gamma_0}^n(\theta)$ in order to emphasize its dependency on the sample size n . With Lemma 1, $\tilde{W}_{\gamma_0}^n(\theta)$ is continuous in θ_0 for large enough n , hence there exists a N_c such that $\tilde{W}_{\gamma_0}^n(\theta)$ is continuous in θ_0 for all $n \geq N_c$. Fix arbitrary $\varepsilon > 0$ and for each $\delta > 0$ define B_δ^n as the set of all θ such that $\|\theta - \theta_0\| < \delta$ with $\|\tilde{W}_{\gamma_0}^n(\theta) - \tilde{W}_{\gamma_0}^n(\theta_0)\| > \varepsilon$. If $\hat{\theta}_m \notin B_\delta^n$ and $\|\tilde{W}_{\gamma_0}^n(\hat{\theta}_m) - \tilde{W}_{\gamma_0}^n(\theta_0)\| > \varepsilon$ then $\|\hat{\theta}_m - \theta_0\| \geq \delta$. As a consequence

$$P\left(\left\|\tilde{W}_{\gamma_0}^n(\hat{\theta}_m) - \tilde{W}_{\gamma_0}^n(\theta_0)\right\| > \varepsilon\right) \leq P\left(\hat{\theta}_m \in B_\delta^n\right) + P\left(\left\|\hat{\theta}_m - \theta_0\right\| \geq \delta\right).$$

$$**P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

The second term on the right hand side converges to 0 as $m \rightarrow \infty$ for every fixed $\delta > 0$, since by assumption $\hat{\theta}_m \xrightarrow{P} \theta_0$ as $m \rightarrow \infty$. For the first term note that $\tilde{W}_{\gamma_0}^n(\theta)$ is continuous in θ_0 for all $n \geq N_c$. Hence by the definition of continuity, $B_\delta^n \rightarrow \emptyset$ as $\delta \rightarrow 0$ and $n \rightarrow \infty$ and therefore $P(\hat{\theta}_m \in B_\delta^n) \rightarrow 0$ as $\delta \rightarrow 0$ and $n \rightarrow \infty$. Overall we get for arbitrary $\varepsilon > 0$

$$P\left(\left\|\tilde{W}_{\gamma_0}^n(\hat{\theta}_m) - \tilde{W}_{\gamma_0}^n(\theta_0)\right\| > \varepsilon\right) \rightarrow 0$$

as $n, m \rightarrow \infty$.

The following establishes the consistency of the estimator $\hat{\gamma}_{m,n}$.

PROOF. Let $\gamma = \gamma_0$. As Assumption 2 satisfies the condition of Lemma 12, we get with the latter

$$\tilde{W}_{\gamma_0}(\hat{\theta}_m) \xrightarrow{P} \tilde{W}_{\gamma_0}(\theta_0)$$

as $n, m \rightarrow \infty$ and it follows

$$-2\left(\tilde{W}_{\gamma_0}(\hat{\theta}_m) - \tilde{W}_{\gamma_0}(\theta_0)\right) = o_p(1). \quad (88)$$

Moreover, Assumption 2 satisfy the condition of Theorem 1 such that for a fixed $s \in \mathbb{R}$

$$-2\tilde{W}_{\gamma_0}(\theta_0) \xrightarrow{d} \chi_q^2 \quad (89)$$

as $n \rightarrow \infty$. Eqs. (88) and (89) satisfy the condition of Theorem 2.7 (iv) in Van der Vaart (1998) and with the latter it follows that

$$-2\tilde{W}_{\gamma_0}(\hat{\theta}_m) \xrightarrow{d} \chi_q^2$$

as $m, n \rightarrow \infty$ which implies that

$$-2\tilde{W}_{\gamma_0}(\hat{\theta}_m) = O_p(1). \quad (90)$$

The latter means that for every $\varepsilon > 0$ there exists a finite M such that

$$\limsup P\left(\left\|-2\tilde{W}_{\gamma_0}(\hat{\theta}_m)\right\| > M\right) < \varepsilon. \quad (91)$$

Since $-2\tilde{W}_{\gamma_0}(\hat{\theta}_m) : \Omega \rightarrow \mathbb{R}_+$, eq. (91) is equivalent to

$$\limsup P\left(-2\tilde{W}_{\gamma_0}(\hat{\theta}_m) > M\right) < \varepsilon. \quad (92)$$

Now let $\gamma \neq \gamma_0$. As Assumption 2 satisfies the condition of Lemma 11, we get for the previous M

$$P\left(-2\tilde{W}_{\gamma}(\hat{\theta}_m) > M\right) \rightarrow 1 \quad (93)$$

as $m, n \rightarrow \infty$ for all $\gamma \neq \gamma_0$. Hence for any $\varepsilon > 0$ there exists a number $N_{2\varepsilon}$ such that

$$P\left(-2\tilde{W}_{\gamma}(\hat{\theta}_m) > M\right) \geq 1 - 2\varepsilon \quad (94)$$

for all $m, n \geq N_{2\varepsilon}$. The latter and eq. (92) imply that the following inequalities holds with probability at least $1 - 2\varepsilon$

$$-2\tilde{W}_{\gamma}(\hat{\theta}_m) > M \geq -2\tilde{W}_{\gamma_0}(\hat{\theta}_m) \quad (95)$$

for all $m, n \geq N_{2\varepsilon}$. According to eq. (95) it follows for $\hat{\gamma}_{m,n} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \left[-2\tilde{W}_\gamma(\hat{\theta}_m) \right]$ that

$$P(\hat{\gamma}_{m,n} = \gamma_0) \geq 1 - 2\varepsilon$$

and hence

$$P(\hat{\gamma}_{m,n} \neq \gamma_0) \leq 2\varepsilon \tag{96}$$

for all $m, n \geq N_{2\varepsilon}$. For the following consider the set $\Gamma_0^C := \Gamma \setminus \{\gamma_0\}$ that contains all $\gamma \in \Gamma$ with $\gamma \neq \gamma_0$. Then eq. (96) can be written as

$$P(\hat{\gamma}_{m,n} \in \Gamma_0^C) \leq 2\varepsilon$$

for all $m, n \geq N_{2\varepsilon}$. Now consider the set $\Gamma_0^\delta := \{\gamma \mid \|\gamma - \gamma_0\| > \delta, \gamma \in \Gamma\}$. As by assumption γ_0 is unique and it follows $\Gamma_0^\delta \subseteq \Gamma_0^C$ for all $\delta > 0$. The latter gives

$$P(\|\hat{\gamma}_{m,n} - \gamma_0\| > \delta) = P(\hat{\gamma}_{m,n} \in \Gamma_0^\delta) \leq P(\hat{\gamma}_{m,n} \in \Gamma_0^C) \leq 2\varepsilon$$

for every $\delta > 0$ and $m, n \geq N_{2\varepsilon}$. Finally, as ε and δ can be chosen arbitrarily close to 0, the asserted convergence in probability is established.